CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 117, No. 6 June 2005

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 6, JUNE 2005

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 6, JUNE 2005

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 6, JUNE 2005

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, 2004

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 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).

Abdala, Caroline
Abel, Sharon
Abraham, Bruce
Abrahams, David
Abramovich, Haim
Adachi, Seiji
Agullo, Joaquim
Ahmad, Bashir
Ahmad, Faiz
Akagi, Masato
Akal, Tuncay
Akamatsu, Tomonari
Akay, Adnan
Alain, Claude
Alam, Sheikh
Albert, Donald
Alexander, Joshua
Alexeev, Alexander
Alipour, Fariborz
Alku, Paavo
Allen, John
Alshits, V.
Alsup, James
Amme, Robert
Anderson, John
Andruski, Jean
Angus, James
Arieh, Yoav
Arnold, Walter
Arnott, William
Arslan, Levant
Arvelo, Juan
Ashokkumar, Muthupandian
Astley, Jeremy
Atchley, Anthony
Atlas, Les
Attenborough, Keith
Au, Whitlow
Audoin, Bertrand
Auer, Edward
Avan, Paul
Averkiou, Michalakis
Bachorowski, Jo-Anne
Backhaus, Scott
Bacon, Sid
Badiey, Mohsen
Badin, Pierre
Baer, Ralph
Baggeroer, Arthur
Bai, Mingsian
Bailly, Gerard
Balant, Anne
Balasubramanium, K.
Ballato, Arthur
Bamber, J.
Barbone, Paul
Bard, Ellen
Barham, Richard
Barnett, David
Barney, Anna
Barron, Michael
Barros, Allan Kardec
Bartel, David
Barton, John
Bass, Henry
Baum, Shari
Baz, Amr
Bazua Duran, Carmen
Bedard, Alfred J.
Bee, Mark
Beecher, Michael D.
Beerends, John
Bell, Kristine
Bellandese, Mary H.
Benjamin, Kim
Benki, Jose
Benoit Bird, Kelly
Benson, Jonathan
Bent, Tessa
Beran, Mark
Berengier, Michel
Berg, Jan
Berger, Elliott
Berglund, Birgitta
Bergmann, Raymond
Berkman, Evan
Berman, David Harry
Bernstein, Leslie
Berryman, James
Berthelot, Yves
Berthommier, Frederic
Bertrand, Olivier
Best, Virginia
Bilsen, Frans
Birdsall, Theodore
Biringen, Sedat
Bishop, Garner
Blacklock, Oliver
Blackstock, David
Blanc-Benon, Philippe
Bloothooft, Gerrit
Blumrich, Reinhard
Blumstein, Dan
Boe, Louis-Jean
Boehnke, Susan
Bohn, Ocke-Schwen
Bonnet, Guy
Bostrom, Anders
Botteldooren, Dick
Boulanger, Philippe
Bowles, Ann
Bowlin, James
Boyce, Suzanne
Brachmanski, Stefan
Bradbury, Jack
Bradley, Charles
Bradley, David
Bradley, John
Braida, Louis
Brand, Thomas
Braunschweig, Thomas
Bravo, Teresa
Brayman, Andrew
Breebaart, Jeroen
Brick, Haike
Brooke, Gary
Brown, Judith
Brown, Lex
Brown, Michael
Brownell, William
Brujan, E. A.
Brungart, Douglas
Buckingham, Michael
Buder, Eugene
Buell, Thomas
Buick, Jim
Bunnell, H. Timothy
Bunton, Kate
Burgess, William
Burkard, Robert
Burkhard, Mahlon
Burnett, Theresa
Burroughs, Courtney
Buss, Emily
Butler, John
Cable, Peter
Calvo, David
Campbell, Nick
Campillo, Michel
Candy, James
Cantrell, John
Capleton, Brian
Carey, William
Cariani, Peter
Carlile, Simon
Carlyon, Robert
Carney, Laurel
Carre, Rene
Carroll, Gerard
Carstensen, Edwin
Cartling, Bo
Castaings, Michel
Catheline, Stefan
Cathignol, Dominique
Cato, Douglas
Cavanaugh, William
Cawley, Peter
Cazals, Yves A. G.
Chadwick, Richard
Chaigne, Antoine
Chalikian, Tigran
Chambers, David
Chambers, James
Chapelon, Jean-Yves
Chapman, David M. F.
Chapman, N. Ross
Charette, Francois
Chatterjee, Monita
Chen, Xucai
Cheng, Anping
Cheng, Arthur
Cheng, Li
Cherng, John
Chester, Pirzanski
Chew, Weng Cho
Cheyne, Harold
Chiaverina, Christopher
Chiu, Ching-Sang
Cholewiak, Roger
Chotiros, Nicholas
Choy, Yat-sze
Christopher, Ted
Chu, Dezhang
Chu, Moody
Church, Charles
Cipolla, Jeffrey
Claesson, Ingvar
Clark, Cathy Ann
Clark, Robert
Clarke, Constance
Clay, Clarence
Cleary, Miranda
Cleveland, Robin
Clorennec, Dominique
Cloutier, Guy
Cobo, Pedro
Cochrane, Norman
Colburn, H. Steven
Collier, Sandra
Collins, Michael D.

 0001-4966/2005/117(6)/3335/6/$20.00

Colonius, Tim
Burnett, David
Colosi, John
Colton, Raymond
Colussi, Augustin
Commander, Kerry
Coombs, Sheryl
Cooper, Nigel
Cornuelle, Bruce
Corsaro, Robert
Covey, Ellen
Cowan, Robert S.
Cox, Trevor
Craig, James
Cranch, Geoffrey
Cray, Benjamin
Cugnet, Boris
Culling, John
Cummings, William
Cuschieri, Joseph
Dahl, Peter
Dai, Huanping
Daigle, Gilles
Dalecki, Diane
D'Alessandro, Christophe
Daley, Tom
Dantzker, Marc
Dargush, G. F.
Darwin, Christopher
Dasgupta, Nilanjan
Dassios, George
Dau, Torsten
Davies, Huw
Davies, Patricia
Davis, Anthony
Davis, Rickie
Dayton, Paul
Deane, Grant
DeBilly, Michel
De Cheveigne, Alain
Declercq, Nico
Deecke, Volker
Defrance, Jerome
Degertekin, Levent
De Groot-Hedlin, Catherine
De Hoop, Adrianus
Delebarre, Christophe
Demoulin, Xavier
Depollier, Claude
De Rosny, Julien
Deshmukh, Om
Desmet, Wim
Destrade, Michel
Devore, Sasha
Dhar, Sumit
Di Benedetto, Maria-Gabriella
Dickey, Joe
DiNapoli, Fred
Doinikov, Alexander
Dolan, David
Donaldson, Gail
Donskoy, Dimitri
Dooling, Robert
Dorman, Michael
Dosso, Stan
Dougherty, Bob
Doval, Boris
Dowling, David
Dreisbach, Laura
Drennan, Ward
Dreyer, Thomas
Dromey, Christopher
Drullman, Rob
Dubno, Judy
Dubus, Bertrand
Duda, Timothy
Duifhuis, Hendrikus
Dunn, Floyd
Durlach, Nathaniel
Durrant, John
Dushaw, Brian
Dvorkin, Jack
Dye, Raymond
Dzieciuch, Matthew
Ebbeson, Gordon
Eddins, David
Edwards, Brent
Eggermont, Jos
Ehrhardt, Matthias
Eisler, Thomas
Elliott, Stephen
Ellis, Daniel
Ellison, William
Elsherbeni, Atef Z.
Emelianov, Stanislav
Enflo, Bengt
Erbe, Christine
Erickson, Molly
Erickson, Robert
Ermann, Michael
Erskine, Fred
Espinoza-Varas, Blas
Evans, Richard
Everbach, Erich
Eversman, Walter
Everstine, Gordon
Ewart, Terry
Fahey, Paul
Fahnline, John
Fain, Sean
Farabee, Theodore
Farassat, Feri
Farina, Angelo
Farmer, David
Farouk, Bakhtier
Fastl, Hugo
Fatemi, Mostafa
Faulkner, Andrew
Fawcett, John
Feleppa, Ernest Joseph
Fellah, Zine el Abiddine
Ferguson, Neil
Ferrara, Katherine
Festen, Joost
Feth, Lawrence
Fidell, Sanford
Filipanits, Frank
Fine, Michael
Finette, Steven
Finlayson, Paul
Finneran, James
Finnveden, Svante
Fishman, Louis
Fitch, W. Tecumseh
Fitzgibbons, Peter
Flatté, Stanley
Fletcher, Neville
Flint, John
Florentine, Mary
Folkins, John
Folkow, Peter
Foote, Kenneth
Ford, John
Fosler-Lussier, Eric
Fourcin, Adrian
Fournier, Daniele
Fox, Robert
Fox, Warren
Frampton, Kenneth
Francis, Alexander
Franck, Andreas
Franco, Horacio
Frankel, Joe
Frankenthal, Shimshon
Franks, John
Franzoni, Linda
Frazer, L. Neil
Frederickson, Carl
Freitag, Lee
Freyman, Richard
Friberg, Anders
Friend, James
Frisk, George
Frizzell, Leon
Fromm, David
Fromme, Paul
Fu, Qian-Jie
Fu, Yibin
Fuchs, Helmut
Fuks, Iosef
Furihata, Kenji
Gabrielson, Thomas
Galland, M. A.
Gallun, Frederick
Galvin, John
Gandour, Jackson
Garai, Massimo
Gardonio, Paolo
Garrelick, Joel
Garrett, Steven
Gaumond, Charles
Gauss, Roger
Gedamke, Jason
Georg, Klump
Georgios, Natsiopoulos
Gerhard-Multhaupt, R.
Gerlotto, Francois
Gerstoft, Peter
Ghitza, Oded
Gibbs, Barry
Gilbert, Kenneth
Gillis, Keith
Gilmore, Grant
Ginsberg, Jerry
Givelberg, Edward
Glegg, Stewart
Glorieux, Christ
Gockel, Hedwig
Godin, Oleg
Goebl, Werner
Goggans, Paul
Goldstein, Louis
Goldsworthy, Ray
Gompf, B.
Goodman, Ralph
Gordon-Salant, Sandra
Gorga, Michael
Goshorn, Edward
Gottlieb, Hans
Gough, Colin
Graham, W.
Granqvist, Svante
Grantham, Wesley
Graziani, Giorgio
Greenberg, Julie
Greene, Charles
Griesinger, David
Grose, John
Grosh, Karl
Grove, Deborah
Guicking, Dieter
Guinan, John
Guion, Susan
Gumerov, Nail A.
Gummer, Anthony
Guo, Jingnan
Guo, Ningqun
Gusev, Vitalyi
Haberman, Michael
Hagiwara, Robert
Haisch, Christoph
Hall, Donald
Hall, Joseph III
Hall, Timothy
Hallworth, Richard
Hammershoi, Dorte
Han, Seon
Handel, Stephen
Hansen, Uwe
Hanson, Helen
Hanyga, Andrzej
Hardin, Jay
Hari, Riitta
Harkin, Antony
Harris, Frances
Harrison, Brian
Harrison, Christopher
Hartmann, William
Hasegawa-Johnson, Mark
Hasselbrink, Charlie
Hastings, Mardi
Hautus, Michael
Hayashi, Takahiro
Hayes, Erin
He, David
Healy, Eric
Heaney, Kevin
Hedrick, Mark
Heffner, Henry
Hefner, Brian
Heinz, Michael
Helffrich, Jerry
Héliea, Thomas
Hellman, Rhona
Hellweg, Robert
Henning, G. Bruce
Henrich, Nathalie
Henyey, Frank
Hermansky, Hynek
Hermes, Dik
Hertrich, Ingo
Herzel, Hanspeter
Heyliger, P. R.
Heyman, Joseph
Hickey, Craig
Hilgenfeldt, Sascha
Hill, Nicolas
Hillenbrand, James
Hinkelman, Laura
Hirose, Sohichi
Hirsekorn, Sigrun
Hixon, Elmer
Hodgkiss, William

Hodgson, Murray
Hoff, Lars
Hofman, Paul
Hole, Lars
Holland, Charles
Holland, Christy
Holland, Stephen
Holliday, D. Vance
Hollien, Harry
Holmberg, Eva
Holmes, Jason
Holt, Lori
Holt, R.
Holzfuss, Joachim
Homm, Anton
Hopgood, James
Horiuchi, Timothy
Hornsby, Benjamin
Horoshenkov, Kirill
Horst, Johannes
Hosten, Bernard
Houde, John
Houser, Dorian
Houtgast, Tammo
Houtsma, Adrianus
Howard, David
Howe, Bruce
Howe, Michael
Hsu, Nelson
Huber, Jessica
Humes, Larry
Hunter, Eric
Hurley, David
Hursky, Paul
Ih, Jeong-Guon
Ihlenburg, Frank
Ilinskii, Yurii
Insana, Michael
Insley, Stephen
Isakov, Victor
Ishimaru, Akira
Iwasa, Kuni
Jackson, Darrell
Jacobs, Laurence
Jacobsen, Finn
Janssen, Thomas
Javel, Eric
Jenison, Rick
Jensen, Finn
Jerger, James
Jesse, Alexandra
Jesteadt, Walt
Jesus, Sergio
Jiang, Jack
Johnson, David
Johnson, Mark
Johnson, Marty
Jones, Christopher
Jones, Jeffery
Jülicher, Frank
Jun, Sun Ah
Juve, Daniel
Kaduchak, Gregory
Kaernbach, Christian
Kahle, Eckhard
Kalveram, Karl
Kammeyer, Karl
Kang, Jian
Kaplunov, Julius
Kappas, Arvid
Kar, T.
Karabutov, Alexander
Kargl, Steven
Karjalainen, Matti
Karniadakis, George
Kastelein, Ronald
Katsnelson, Boris
Katz, Brian
Katz, Larry
Katz, William
Kaufman, Jonathan
Keefe, Douglas
Keenan, Ruth
Keith, William
Kenbu, Teramoto
Kent, Raymond
Keolian, Robert
Kergomard, Jean
Kewley-Port, Diane
Khismatullin, Damir
Kidd, Gary
Kiefte, Michael
Kim Jaehwan
Kim, Jay
Kim, Kang
Kim, Sang-Myeong
King, John R.
King, Simon
Kingston, John
Kirby, Raymond
Kirkeby, Ole
Kirkpatrick, Sean
Kitahara, Mafuyu
Klaboe, Ronny
Klapuri, Anssi
Klis, S. F.
Kluender, Keith
Knobles, David
Kob, Malte
Kobayashi, Michiaki
Koenig, Laura
Koessl, Manfred
Kohler, Klaus
Kole, Sebastian
Kolios, Michael
Kollmeier, Birger
Koopmann, Gary
Korman, Murray
Korneliussen, Rolf
Kortekaas, Reinier
Kosinski, John A.
Kossl, Manfred
Kozick, Richard
Krause, Jean
Kreiman, Jody
Kremkau, Frederick
Kressmann, Reinhardt
Krishna, Suresh
Krishnamurthy, Ashok
Krishnan, Ananthanarayan
Krishnaswamy, Sridhar
Kristiansen, Ulf
Kropp, Wolfgang
Krutitskii, Pavel
Krylov, Victor
Kuang, Z. B.
Kumon, Ronald
Kundu, Tribikram
Kuperman, William
Kuttruff, K. Heinrich
Lacefield, James
Lakatos, Stephen
Lambert, Jacques
Lamberti, Nicola
Lammers, Marc
Lane, Courtney
Lane, Steve
Lanza di Scalea, Francesco
Large, Edward
Larraza, Andres
Larson, Gregg
Lauchle, Gerald
Lauri, Olivier
Lauriks, Walter
Lavandier, Catherine
Lee, Chin-Hui
Lee, Hua
Lee, Jang Moo
Lee, Joon-Hyun
Lee, Lin-Shan
Lehman, Sean
Leishman, Timothy
Lentz, Jennifer
LePage, Kevin
Lethiecq, Marc
Letowski, Tomasz
Leurer, Klaus
Levy, Moises
Lewin, Peter
Li, Bengang
Li, Kai Ming
Licht, Torben R.
Lieberman, Philip
Lieuwen, Tim
Lim, Teik
Lineton, Ben
Ling, Ju
Lingevitch, Joseph
Lipkens, Bart
Liss, Julie
Litovsky, Ruth
Litt, Mitchell
Liu, Donald
Liu, Thomas
Lizzi, Frederic
Lobel, Phil
Lobkis, Oleg
Loeppert, Peter
Loizou, Philipos
Long, Glenis
Lorenzi, Christian
Lotto, Andrew
Love, Richard
Lowe, Michael
Lubman, David
Lucks Mendel, Lisa
Lueptow, Richard
Lukashkin, Andrei
Lunde, Per
Lutfi, Robert
Lynch, James
Lyon, Richard
Lyons, Anthony
Ma, Yuanliang
Macken, William J.
MacMillan, Neil
Madanshetty, Sameer
Madsen, Peter
Maeda, Shinji
Maekawa, Zyun iti
Maev, Roman
Maison, Stephane
Maki, Katuhiro
Makris, Nicholas C.
Maltsev, Nick
Manley, Geoffrey
Mann, David
Marengo, Edwin
Marshall, Bill
Marshall, Lynne
Marston, Philip
Martens, William
Martin, Brett
Martin, Glen
Martin, Siderius
Martinez, Rudolph
Marx, David
Massaro, Dominic
Mast, T. Douglas
Matveev, Konstantin
Mauermann, Manfred
Max, Michael
Maynard, Julian
Maysenhoelder, W.
McAdams, Stephen
McConnell, James
McCoy, John
McDaniel, James
McDermott, Hugh
McDonald, Bryant
McDonald, Mark
McGehee, Duncan
McGowan, Richard
McInerny, Sally
McKay, Colette
Meddis, Ray
Medwin, Herman
Meegan, George
Melamed, Timor
Mellinger, David
Mellody, Maureen
Mettin, Robert
Michael, Michael
Michaels, Jennifer
Michalopoulou, Zoi-Heleni
Micheyl, Christophe
Middlebrooks, John
Miedema, Henk
Migliori, Albert
Mignerey, Peter
Miles, Ronald
Miller, Charles
Miller, Douglas
Miller, James
Mills, David
Milner, Ben
Ming, Ji
Misun, Vojtech
Mitri, Farid
Mitson, Ronald
Miyara, Federico
Mobley, Joel
Moehl, Bertel
Moeser, Michael
Moffett, Mark
Mohl, Bertel
Moldover, Michael
Molis, Michelle
Monaghan, Alex
Monahan, Edward
Mongeau, Luc
Montanaro, A.
Montero de Espinosa, Francisco

Moon, Won-kyu
Moore, Brian
Moore, Patrick
Moorhouse, Andy
Moreton, Elliott
Morioka, Ikuharu
Morro, Angelo
Morse, Scot
Morton, Eugene
Mote, C.
Moulin, Emmanuel
Mozhaev, Vladimir
Mozurkewich, George
Muehleisen, Ralph
Mueller, Tobias
Muesch, Hannes
Muir, Thomas
Mullennix, John
Munhall, Kevin
Munjal, M.
Murphy, Peter
Murray, Bruce
Musicant, Alan
Nabelek, Anna
Nachtigall, Paul
Nagem, Raymond
Naghshineh, Koorosh
Nagy, Peter
Nakamura, Isao
Naylor, Graham
Nearey, Terrance
Neely, Stephen
Neff, Donna
Nelson, David
Nelson, Peggy
Newman, Rochelle
Ng, Manwa
Niethammer, Marc
Nightingale, K. R.
Nilsson, Michael
Nittrouer, Susan
Nolle, Alfred
Nolte, Bodo
Nordholm, Sven
Norris, David
Norton, Guy
Norton, Stephen
Novarini, Jorge
Nowacek, Douglas
Nusbaum, Howard
Nyborg, Wesley
Nye, Patrick
Oba, Roger
O'Brien, William, Jr.
Ochmann, Martin
Oded, Ghitza
Odom, Robert
Ogam, Erick
Ogi, Hirotsugu
Ohde, Ralph
Ohkawa, Keiichi
Ohl, Claus-Dieter
Ohno, Masahiro
Okal, Emile
Oldham, David
Olive, Sean
Oller, Kimbrough
Olson, Elizabeth
Onuu, Michael
Orduna-Bustamante, Felipe
Orris, Gregory
Ostashev, Vladimir
Ostendorf, Mari
Ostoja, Martin
Ostrovsky, Lev
Oswald, Julie
Ouni, Slim
Owren, Michael
Oxenham, Andrew
Paek, Insu
Palmer, David
Palmer, Stuart
Paltauf, Guenther
Pan, Jie
Pandit, Aniruddha
Panneton, Raymond
Pantle, Iris
Papanicolaou, George C.
Pardo, Jennifer
Pardue, Sally
Parkins, John
Parsa, Vijay
Pascal, Jean-Claude
Pasqualini, Donatella
Patat, Frederic
Patterson, Roy
Pavan, Gianni
Pavlovic, Chaslav
Payton, Karen
Pearsons, Karl
Peat, Keith
Pecorari, Claudio
Pedersen, Michael
Pell, Marc
Pelorson, Xavier
Peplow, Andrew
Petculescu, Andi
Peters, Robert
Petropoulos, Peter
Pfingst, Bryan
Phillips, Dennis
Piacsek, Andrew
Picaut, Judicaël
Pierce, Allan
Pierrehumbert, Janet
Pinsky, Peter
Pittman, Andrea
Plack, Christopher
Plona, Thomas
Plotkin, Kenneth
Poese, Matthew
Pollard, Howard
Poncelet, Olivier
Ponton, Curtis
Popel, Aleksander
Port, Robert
Porter, Tyrone
Poterek, Thomas
Potty, Gopu
Prada, Claire
Prasad, Ajay
Prasanna, S. R. M.
Preisig, James
Premus, Vincent
Pressnitzer, Daniel
Preumont, Andre
Prieve, Beth
Prosperetti, Andrea
Przekop, Adam
Puria, Sunil
Qu, Jianmin
Quené, Hugo
Quinn, Robert
Rabinkin, Daniel
Rackl, Robert
Rafaely, Boaz
Rajan, Subrahaniam
Rajapan, Dhilsha
Raju, P.
Rakerd, Brad
Ralston, James M.
Ramakrishnan, Ramani
Ramsay, Gordon
Rasmussen, Karsten
Rauschecker, Josef
Ravazzani, Paolo
Ravier, Philippe
Razansky, Daniel
Recio, Alberto
Redford, Melissa
Remenieras, Jean-Pierre
Ren, Tianying
Rendall, Drew
Renshaw, Anthony
Richards, Virginia
Richardson, Bernard
Richardson, Michael
Rickard, Scott
Rienstra, Sjoerd
Rindel, Jens
Roberts, Brian
Robinson, Harold
Rocchesso, Davide
Rodriguez, Orlando
Rogers, Catherine
Rogers, Peter
Roh, Yongrae
Romeo, Maurizio
Rose, Francis
Rose, Joseph
Rosen, Stuart
Rosenberg, Allan
Rosenthal, Jordan
Rosowski, John
Rossing, Thomas
Rouseff, Daniel
Roux, Philippe
Royer, Daniel
Royston, Thomas
Rumsey, Francis
Sabatier, James
Sabo, Diane L.
Sabra, Karim
Sakagami, Kimihiro
Salt, Alec
Sanghvi, Narendra
Sapatinas, Theofanis
Sapozhnikov, Oleg
Sarabandi, Kamal
Sarkar, Kausik
Sarkissian, Angie
Saurenman, Hugh
Sboros, Vassilis
Scalerandi, Marco
Scales, John
Scharenborg, Odette
Scharf, Bertram
Scharstein, Robert
Schellenberg, Glenn
Scherer, Ronald
Schlauch, Robert
Schmidt, Henrik
Schmitt, Denis
Schneider, Hans Georg
Schobben, Daniel
Schoentgen, Jean
Schomer, Paul
Schröder, Wolfgang
Schuhmacher, Andreas
Schumacher, Robert
Schusterman, Ronald
Schwartz, Jean-Luc
Schwartz, Joshua
Selamet, Ahmet
Semenova, Tatiana
Sen, Mrinal
Sethares, William
Sgard, Franck
Shackleton, Trevor
Shadle, Christine
Shamma, Shihab
Shang, Er-Chang
Shannon, Robert
Shapiro, Serge
Sharma, Anu
Sheft, Stanley
Shen, Steve
Shepard, W.
Shera, Christopher
Sherer, Scott E.
Shimizu, Yasushi
Shore, Susan
Shorter, Philip
Shriberg, Elizabeth
Shrivastav, Rahul
Shung, K. Kirk
Shuyu, Lin
Shynk, John
Siebein, Gary
Siegel, Jonathan
Simmen, Jeff
Simmons, Andrea
Simmons, James
Simonetti, Francesco
Simpson, Harry
Simpson, Roger
Sinha, Bikash
Sinnott, Joan
Skarsoulis, Emmanuel
Skowronski, Mark
Skvor, Zdenek
Slaney, Malcolm
Smeulders, David
Smith, Bart
Smith, David Eric
Smith, Jerry
Smith, Julius
Smith, Leslie
Smith, Suzanne
Snow, David
Snyder, Scott
Soize, Christian
Solodov, Igor
Sommerfeldt, Scott
Sondhi, Mohan
Song, Hee-Chun
Sorokin, Victor
Southall, Brandon
Sparrow, Victor
Spector, Alexander
Sperry, Brian
Spicer, Jim
Spiesberger, John
Spindel, Robert

Spisar, Monica
Spoor, Philip
Sproat, Richard
Stanic, Steve
Stathopoulos, Elaine
Steele, Charles
Stellmack, Mark
Stelmachowicz, Pat
Stickney, Ginger
Stinson, Michael
Stoel-Gammon, Carol
Stojanovic, Milica
Stokes, Ann
Storey, Brian
Stork, Wilhelm
Story, Brad
Stotts, Steven
Stoughton, John W.
Stramaglia, Sebastiano
Strange, Winifred
Strasberg, Murray
Strickland, Elizabeth
Strube, Hans Werner
Stuart, John
Studebaker, Gerald
Sturm, Frederic
Stylianou, Yannis
Succi, George
Sujith, R.
Sullivan, Edmund
Sum, K. S.
Summers, Ian
Summers, Jason
Summers, Van
Sun, Jian-Qiao
Sundberg, Johan
Sung, Chia-Chi
Surprenant, Aimee
Sussman, Elyse
Sussman, Joan
Sutin, Alexander
Svensson, Peter
Svirsky, Mario
Swanson, David
Swift, Gregory
Syrdal, Ann
Szabo, Thomas
Taflove, Allen
Tajima, Keiichi
Takagi, Kenshiro
Talmadge, Carrick
Talmant, Marylin
Tanter, Mickael
Tarnow, Viggo
Taroudakis, Michael
Tavakkoli, Jahan
Telschow, Ken
TenCate, James
Ter Haar, Gail
Terhune, Jack
Terray, Eugene
Teutsch, Heinz
Theunissen, Frederic
Thibodeau, Linda
Thode, Aaron
Thomenius, Kai Erik
Thompson, Lonny
Thomsen, Frank
Thomson, David
Thomson, Ron
Thorsos, Eric Ivan
Thorsson, Pontus
Tindle, Christopher
Tito, Frank
Titze, Ingo
Toiviainen, Petri
Tomilina, Tatiana
Torre, Peter
Tourin, Arnaud
Touzé, Cyril
Tracey, Brian
Trahiotis, Constantine
Tricas, Timothy
Tsingos, Nicolas
Tsogka, Chrysoula
Tuller, Betty
Tun, Pat
Turgut, Altan
Turner, Christopher
Turner, Joseph
Tyack, Peter
Tzanetakis, George
Uchanski, Rosalie
Unger, Evan
Unoki, Masashi
Uscinski, Barry
Vainio, Martti
Valdivia, Nicolas
Valimaki, Vesa
Vallidis, Nicholas
VanBuren, Arnie
Van Compernolle, Dirk
Van de Par, Steven
Van den Doel, Kees
Van der Heijden, Marcel
Van Dinther, Ralph
Van Kamp, Irene
Van Parijs, Sofie
Van Son, Rob
Van Tasell, Dianne
Van Wijngaarden, Sander
Varadan, Vasundara
Varghese, Tomy
Vassilakis, Pantelis
Vatikiotis-Bateson, Eric
Vera, Michael
Verkeyn, Andy
Vermeir, Gerrit
Vidmar, Paul
Vipperman, Jeffrey
Virovlyansky, Anatoly
Vlahopoulos, Nickolas
Vliegen, Joyce
Von Wagner, Utz
Vorobyov, Sergiy
Voronovich, Alexander
Vos, Joos
Voss, Susan
Waag, Robert
Wada, Hiroshi
Wagstaff, Ronald
Wahlberg, Magnus
Wakeland, Ray
Walhberg, Magnus
Walker, Bruce
Walsh, Timothy
Walthelm, Axel
Walton, Joseph
Wang, Can-yun
Wang, Ding
Wang, Emily
Wang, Jack
Wang, Jiqing
Wang, Lugen
Wapenaar, Cornelis P.
Ward, Darren
Ward, Lawrence
Warnock, Alfred
Warren, Richard
Warusfel, Olivier
Waters, Kendall
Watkins, Anthony
Waxler, Roger
Weaver, Richard
Weiland, Nathan
Weinreich, Gabriel
Weisenberger, Janet
Weismer, Gary
Wells, Bill
Werner, Lynne
West, James
Westbury, John
Wester, Eric
Westwood, Evan
White, Michael
Wickert, Jonathan
Wiegrebe, Lutz
Wightman, Fred
Wilen, Larry
Wilkens, Roy
Williams, Earl
Williams, Kevin
Wilson, D. Keith
Wilson, Preston
Wirgin, Armand
Withnell, Robert
Witten, Alan
Wojtczak, Magdalena
Wolfe, Patrick
Wolfson, Michael
Woloszyn, Philippe
Wong, George
Wong, Patrick
Woodhouse, Jim
Woods, William
Worcester, Peter
Woszczyk, Wieslaw
Wouters, Jan
Wright, Andrew
Wright, Beverly
Wright, Selwyn
Wright, William
Wu, Chung-Hsien
Wu, Hwai-Chung
Wu, Junru
Wu, Kuangcheng
Wu, Sean
Wu, Ting-Wen
Wu, Tsung-Tsong
Wurmser, Daniel
Xiang, Ning
Xu, Yi
Yan, Hong
Yang, Tsih
Yang, Xinmai
Yang, Yao-Joe
Yano, Takeru
Yevick, David
Yoshikawa, Shigeru
Yoshioka, Hirohide
Yost, William
Zabolotskaya, Evgenia
Zacharov, Nick
Zagrai, Andrei
Zahorian, Stephen
Zahorik, Pavel
Zahui, Marcellin
Zaitsev, V.
Zakharov, Dmitry
Zeng, Fan-Gang
Zeqiri, Bajram
Zhang, Kechen
Zhao, Xiaoliang
Zheng, Yibing
Zheng, Z. Charlie
Zhong, Pei
Zhou, Ji-Xun
Ziegler, Wolfram
Zimmer, Walter
Ziomek, Lawrence
Zuckewar, Allan
Zurek, Patrick

USA Meetings Calendar

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

2005

18–22 July	17th International Symposium on Nonlinear Acoustics, State College, PA [Anthony Atchley, The Pennsylvania State University, 217 Applied Research Lab Building, University Park PA 16802; Tel.: 814-865-6364; E-mail: ISNA17@outreach.psu.edu; WWW: http://www.outreach.psu.edu/c&i/isna17/].
17–21 October	150th Meeting joint with Noise-Con 2005, Minneapolis, MN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Te.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
27–29 October	5th International Symposium on Therapeutic Ultrasound, Boston MA [E-mail: info@istu2005.org; WWW: www.istu2005.org].

2008

28 July–1 Aug	9th International Congress on Noise as a Public Health

Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609; Tel. 860-572-0680; E-mail: icben2008@att.net; WWW: www.icben.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

ACOUSTICAL NEWS—INTERNATIONAL

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

International Meetings Calendar

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

June 2005

1–3 **1st International Symposium on Advanced Technology of Vibration and Sound**, Hiroshima, Japan (Web: dezima.ike.tottori-u.ac.jp/vstech2005).

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

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

27–29 **Managing Uncertainties in Noise Measurements and Prediction**, Le Mans, France (Web: www.uncertainty-noise.org).

28–1 **International Conference on Underwater Acoustic Measurements: Technologies and Results**, Heraklion, Crete, Greece (Web: UAmeasurements2005.iacm.forth.gr).

July 2005

4–8 **Turkish International Conference on Acoustics 2005: New Concepts for Harbor Protection, Littoral Security, and Underwater Acoustic Communications**, Istanbul, Turkey (Web: www.tica05.org/tica05)

11–14 **12th International Congress on Sound and Vibration**, Lisbon, Portugal (Web: www.icsv12.ist.utl.pt).

August 2005

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

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

28–1 **World Congress on Ultrasonics Merged with Ultrasonic International (WCU/UI'05)**, Beijing, China (Secretariat of WCU 2005, Institute of Acoustics, Chinese Academy of Sciences, P.O. Box 2712 Beijing, 100080 China; Fax: +86 10 62553898; Web: www.ioa.ac.cn/wcu-ui-05).

31–3 ***6th Pan European Voice Conference**, London, UK (Web: www.pevoc6.com/home htm).

September 2005

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

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

18–21 **IEEE International Ultrasonics Symposium**, Rotterdam, The Netherlands (Web: www.ieee-uffc.org).

20–22 **International Symposium on Environmental Vibrations**, Okayama, Japan (Web: isev2005.civil.okayama-u.ac.jp).

27–29 **Autumn Meeting of the Acoustical Society of Japan**, Sendai, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).

October 2005

12–14 **Acoustics Week in Canada**, London, Ontario, Canada (Web: caa-aca.ca).

17–18 **Wind Turbine Noise: Perspectives for Control**, Berlin, Germany (G. Leventhall, 150 Craddocks Avenue, Ashtead Surry KT21 1NL UK; Fax: +44 1372 273 406; Web: www.windturbinenoise2005.org).

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

25–26 **Autumn Conference 2005 of the UK Institute of Acoustics**, Oxford, UK (Web: www.ioa.org.uk).

27–28 ***Autumn Meeting of the Acoustical Society of Switzerland**, Aarau, Switzerland (Web: www.sga-ssa.ch).

November 2005

4–5 **Reproduced Sound 21**, Oxford, UK (Web: www.ioa.org.uk).

9–11 **Australian Acoustical Society Conference on "Acoustics in a Changing Environment,"** Busselton, WA, Australia (Web: www.acoustics.asn.au/divisions/2005-conference.shtml).

14–18 ***XVI Session of the Russian Acoustical Society**, Moscow, Russia (Web: www.akin.ru).

December 2005

7–9 ***Symposium on the Acoustics of Poro-Elastic Materials**, Lyon, France (Fax: +33 4 72 04 70 41; Web: v0.intellagence.eu.com/sapem2005).

January 2006

5–7 ***First International Conference on Marine Hydrodynamics**, Visakhapatnam, India (V. B. Rao, Naval Science & Technological Laboratory, Vigyan Nagar, Visakhapatnam—530 027, India; Web: www.mahy2006.com).

May 2006

15–19 **IEEE International Conference on Acoustics, Speech, and Signal Processing**, Toulouse, France (Web: icassp2006.org).

June 2006

5–6 **6th European Conference on Noise Control (EURONOISE2006)**, Tampere, Finland (Fax: +358 9 7206 4711; Web: www.acoustics.hut.fi/asf).

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

July 2006

3–7 **13th International Congress on Sound and Vibration (ICSV13)**, Vienna, Austria (Web: info.tuwienac.at/icsv13).

17–19 **9th International Conference on Recent Advances in Structural Dynamics**, Southampton, UK (Web: www.isvr.soton.ac.uk/sd2006/index.htm).

 0001-4966/2005/117(6)/3341/2/$20.00

September 2006

13–15 **Autumn Meeting of the Acoustical Society of Japan**, Kanazawa, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).

July 2007

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

August 2007

27–31 **Interspeech 2007**, Antwerp, Belgium (e-mail: conf@isca-speech.org).

September 2007

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

9–12 **ICA Satellite Symposium on Musical Acoustics (ISMA2007)**, Barcelona, Spain (SEA, Serano 144, 28006 Madrid, Spain; Web: www.ica2007madrid.org).

June 2008

23–27 **Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**, Paris, France (E-mail: phillipe.blanc-benon@ec-lyon.fr).

July 2008

28–1 **9th International Congress on Noise as a Public Health Problem**, Mashantucket, Pequot Tribal Nation (ICBEN 9, P.O. Box 1609, Groton CT 06340-1609, USA: Web: www.icben.org).

Preliminary Announcement

August 2010

TBA **20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (Web: www.acoustics.asn.au).

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).]*

Acoustic Absorbers and Diffusers, Theory, design and application

T. J. Cox and P. D'Antonio

Spon Press, London, 2004. 405 pp. Price: $159.95 (hardcover).
ISBN: 0-415-29649-8

After Marshall, Hyde, and Barron first successfully installed number-theoretic (''Schroeder'') diffusors in Wellington Town Hall in New Zealand, Peter D'Antonio was the first entrepreneur to make reflection phase gratings (RPGs) based on quadratic residues widely available. The present volume is the gist of his and his collaborators' experiences with these structures and their cousins in the field of architectural acoustics during the last quarter century.

After a general introduction, separate chapters focus on basic principles of absorbers and diffusors and the measurement of their properties. Porous and resonant absorbers are discussed in detail, as are random geometric and number-theoretic diffusors both in one and two dimensions. Of particular interest is a final chapter on active absorption and diffusion, as well as hybrid (active-passive) systems.

The book is profusely and exceedingly well illustrated. There are numerous helpful references and several appendices on absorption coefficients (from draperies to wooden pews), seven MATLAB programs, extensive diffusion-coefficient tables (from semicylinders and triangles to quadratic-residue and primitive-root-diffusors).

Diffusor applications to roadside noise barriers, ''street canyons,'' as well as recording studios and music practice rooms are likewise treated.

Of particular interest are the authors' fractal diffusors (''diffractals''), which embed high-frequency diffusors within a low-frequency diffusor. I was also intrigued by their two-dimensional hybrid diffusors, which make use of perforated masks of binary maximum-length sequences whose period length can be factored. Thus, for example, $2^{10}-1=1023=31\times 33$. (Such masks have also been used in x-ray astronomy for imaging distant x-ray emitting stars. There seems to be no end to the useful applications of number theory to practical problems).

The book can be highly recommended to workers in the field, as well as scientists interested in a broad range of problems and solutions for acoustical absorbers and diffusors for all kinds of waves (sonar, radar, light).

MANFRED SCHROEDER
University of Goettingen
Drittes Physikalisches Institut
37073 Goettingen, Germany

 0001-4966/2005/117(6)/3343/1/$20.00

OBITUARIES

Robert Sydney Gales • 1914–2004

Robert Sydney Gales, a Fellow of the Society, died in his home in San Diego on May 31, 2004. Mr. Gales was a former President of the Society, a leader in underwater sound and human factors research for the Navy, and was a consultant in architectural acoustics.

Robert Gales was born in Boston on December 12, 1914. Shortly thereafter, his family moved to California. In 1938, he was awarded a BA in Physics from the University of California at Los Angeles (UCLA), and in 1942 he received the MA in Applied Physics from the same institution. During 1938–1942 he worked at UCLA, first on hearing and speech intelligibility under the guidance of Norman Watson, and later on the design of ear protectors for high noise level areas under the guidance of Vern Knudsen and Norman Watson.

From 1942 to 1948, he was an Associate Physicist at the War Research Division of the University of California in San Diego, subsequently renamed the Marine Physical Laboratory (UCSD) where he conducted research on ambient sea noise and aural detection of ship and submarine sounds. He then joined the Navy Electronics Laboratory in San Diego, later renamed the Naval Ocean Systems Center (NOSC). He subsequently became head of the Acoustics, Behavior, and Communication Division of the Biosystems Research Department, and then head of the Airborne Acoustics Branch. His research activities included detection and classification of sonar signals, studies of ambient sea noise, speech communication in the presence of noise, and the effects of airborne noise on personnel. During this period he authored or coauthored over 60 papers, reports, and patents, and contributed to many books on underwater sound, noise control, and human engineering. He retired from NOSC in 1980.

Mr. Gales joined the Acoustical Society of American (ASA) in 1940, and was elected a Fellow in 1950. He served on the Executive Council from 1965 to 1968, and Vice President in 1972–1973, and President in 1975–1976. He participated in many ASA committees, including the Coordinating Committee on Environmental Acoustics. As part of this activity, he chaired the San Diego Workshop on the Interaction between Man-made Noise and Vibrations and Arctic Marine Wildlife. A result of this conference was a research plan to be followed by Federal, State and Native entities for developing solutions to noise problems. Mr. Gales was a member of the American National Standards Institute (ANSI), the Acoustical Standards Board, and the three ANSI acoustics standards committees managed by the ASA (physical acoustics, psychological and physiological acoustics, and noise). He was a member of the National Research Council Committee on Hearing and Bioacoustics from 1960 to 1980.

Mr. Gales was a long-term participant in the San Diego Chapter of the ASA, which he chaired in the early 1960s. He also served on the Noise Control Board of the County of San Diego. His acoustical consulting practice, in association with Robert Young, involved over one hundred significant and varied projects.

He is survived by his wife of 61 years, Dede, their three children, and three grandchildren.

KENNETH M. ELDRED
WILLIAM C. CUMMINGS

John A. Curtis • 1926–2004

John A. (Jack) Curtis, a consultant in architectural acoustics and long-term member of the Society died on December 5, 2004 in his home in Sandwich, Massachusetts. He was born in 1926 in Harbin, the capitol of Manchukuo, in northeast China, where his father was responsible for operations of the National City Bank of New York, and he grew up in Tokyo, where he attended the American School.

Mr. Curtis's early professional training was in architecture; he received an undergraduate degree summa cum laude from Princeton in 1950 and the degree of Master in Architecture from the Massachusetts Institute of Technology in 1953. He had a Travelling Fellowship with the architectural firm, Skidmore, Owings and Merrill, and he also worked with the architectural firm of Iversen van Sitteren in Malaysia.

He joined Bolt, Beranek, and Newman (BBN) in 1957 with the original intent of only staying a short period, so that he could learn more about architectural acoustics from the recognized authorities, such as Robert Newman and William Cavanaugh, who were there at the time. But he liked the BBN environment and the colleagues there so well that he stayed for 29 years, until he retired in 1986. Positions he held at BBN included those of senior acoustical consultant and Director of the Division of Architectural Acoustics. Over these years, he provided consulting services to hundreds of architectural clients, and he strongly influenced the professional development of many well-known workers in architectural acoustics, including Robert Newman, Theodore Schultz, Robert Hoover, William Cavanaugh, Thomas Horrall, Carl Rosenberg, Rein Pirn, Parker Hirtle, Jacek Figwer, and David Kaye. The total number of projects, covering a wide variety of topics, is estimated to be in the thousands. Architectural firms with which he collaborated include The Architects Collaborative, Cambridge Seven, Hugh Stubbins, Benjamin Thompson, Kallmann McKinnell and Wood, and the Architectural Resources of Cambridge. His projects include the AIA Headquarters (The Architects Collaborative), the National Technical Institute for the Deaf at the Rochester Institute of Technology (Hugh Stubbins Associates), IBM Corporate Technical Institute (RTKL), and the Hynes Auditorium and Convention Center (Kallmann McKinnell & Wood).

His principal contributions to the field include the application of acoustical criteria and the use of electronic sound masking in office buildings to enhance acoustical privacy. He was a leader in the use of Speech Privacy Analysis methodology and in the development of computer software that incorporated such methodology. He developed guidelines for office privacy that continue in use today. His writings and presentations in this area include ''Simulated Background Noise to Improve Acoustical Privacy in Buildings,'' presented at the Acoustical Society of America meeting in NYC in 1964 and subsequently published as a report by BBN. Other publications that had substantial impact on the field of office acoustics include ''Noise in the Office'' (National Office Products Association, 1969), ''Taking the Guesswork Out of Office Acoustics'' (BBN, 1979), and ''Improving Office Acoustics'' (Corporate Design Magazine, 1983).

Another principal accomplishment, with Robert Hoover and Robert Jones, was in the development, design, and evaluation of the air conditioning and distribution systems and related noise control methods for concert halls. The work was reported in a presentation, ''Noise Control for Air Conditioning Systems Installed in the Philadelphia Academy of Music and Boston Symphony Hall,'' at the ASA meeting in April 1975. Another activity, with William Cavanaugh, Parker Hirtle, and William Watters, was the development of sound isolation requirements and criteria for university study and dormitory facilities. Results were reported at May 1963 and April 1970 meetings. There was also work on criteria for acoustical design of studio spaces used to produce educational television programs, with the subsequent publication of an article ''Coping with Acoustical Headaches''

 0001-4966/2005/117(6)/3344/2/$20.00

(Educational Television, December 1968). Another influential article was ''Acoustical Considerations in Integrated Ceilings'' (Illuminating Engineering, August 1961). He was prominent among those members of the BBN staff who were honored in 1984 by The American Institute of Architects as ''pioneering acoustical consultants who in thirty-five years have almost single-handedly invented an entire profession by creating an awareness of acoustical considerations in design and by integrating technical solutions based on scientific principles with architectural and artistic concepts.''

After his retirement from BBN, Mr. Curtis was a supporter and member of the Advisory Committee for the Robert Bradford Newman Student Award Fund, established in 1984 and now administered by the Technical Committee on Architectural Acoustics of the Acoustical Society of America. He wrote the annual newsletter for the fund for many years, which gave more than 150 Newman Medals to students at more than 35 universities around the world, as well as awards of Theodore John Schultz Grants supporting research and development of new teaching methods in architectural acoustics.

ERIC W. WOOD
THOMAS R. HORRALL
CARL J. ROSENBERG
ROBERT M. HOOVER
WILLIAM J. CAVANAUGH

Donald Redfield Griffin • 1915–2003

The Society notes the passing of Donald Redfield Griffin, a noted biologist, who, with Robert Galambos, discovered that bats use biological sonar to perceive their surroundings. Griffin's work is extensively cited in this Journal, and he is generally recognized as one of the major contributors to the development of the emerging field of animal bioacoustics. He never joined our Society, probably because at the time when Griffin was doing his most significant work, the Society was not as actively involved in this aspect of acoustics as it is now. Two special sessions in his memory were held at the New York City meeting in May 2004. The sessions were sponsored by the Animal Bioacoustics Technical Committee, with the title ''Natural acoustic behavior of animals: Session in memory of Donald R. Griffin'' [J. Acoust. Soc. Am. **115**(5), Part 2 (2004)]. The abstracts from those sessions give a strong affirmation of the impact of Griffin's work on the research of many members of this Society.

Griffin died in November 2003, in Lexington, MA, at the age of 88. He was born on August 3, 1915, in Southampton, NY. He attended Tabor Academy in Marion, MA, and was a student at Harvard University, receiving his B.S. (1938), M.A. (1940), and Ph.D. (1942) degrees there. While an undergraduate at Harvard, he banded bats to follow their seasonal migrations, and in the course of this work became familiar with the long-standing question of the sensory basis of their orientation at night. It was then that he collaborated with Robert Galambos to establish its acoustic nature. His doctoral work on migration by birds was supervised by Karl Lashley, a pioneer in behavioral research on brain function. From 1942 to 1945 he was a Research Associate at Harvard, where he worked on war-related projects with S. S. Stevens. Following World War II, he went to the faculty of Cornell University as Assistant Professor of Zoology (1946–47), Associate Professor of Zoology (1947–1952), and Professor of Zoology (1952–1953). In 1953, he returned to Harvard as Professor of Zoology and served as Chairman of the Department of Biology from 1962 to 1965. During this time at Harvard, he wrote the influential book, *Listening in the Dark* (1958), which recounted the discovery of echolocation and what had been learned in the early years of research thereafter. In 1965, Donald Griffin moved to the Rockefeller University in New York City. Griffin's laboratories were variously at Theobald Smith Hall, at the New York Zoological Society in the Bronx, and later at the field research center in Millbrook, NY. When Griffin retired from Rockefeller in 1986, he spent a year at Princeton University and then returned to Harvard, where he worked at the Concord Field Station and occasionally taught undergraduates. In this final period of his life he continued his experimental work on bats, birds, and beavers.

Donald Griffin was awarded honorary Doctor of Science degrees from Ripon College in 1966 and from Eberhard-Karls Universität (Tübingen) in 1988. He was a member of the National Academy of Sciences, American Academy of Arts and Sciences, American Philosophical Society, Animal Behavior Society, American Society of Zoologists, Ecological Society of America, and the American Physiological Society. He served as a trustee of the Rockefeller University and was President of the Harry Frank Guggenheim Foundation from 1979 to 1983. For many years he had an association with the Woods Hole Oceanographic Institution, where he was elected a member of the Corporation in 1964.

For a fuller account of Griffin's life and for a list of some of his more significant publications, one may consult the biographical memoir published by the National Academies Press, currently viewable online at the site http://www.nap.edu/readingroom/books/biomems/dgriffin.html

JAMES A. SIMMONS

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.

Reviews of Acoustical Patents

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Univ. of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,846,365

43.25.Yw METHOD AND APPARATUS FOR ACOUSTIC SUPPRESSION OF CAVITATION

Sameer I. Madanshetty, Manhattan, Kansas
25 January 2005 (Class 134/1); filed 30 May 2002

The premise of this patent is that a cavitation-proof surface can be created within a fluid medium by biasing the acoustic output of a nearby transducer so that only compressive, or positive pressure, waves are radiated. How this is accomplished is not abundantly clear. One suggested method seems to be to excite a transducer with a biased electrical input so that there is no negative input signal and, by supposition, no negative acoustic output. However, since there is by definition no dc acoustic signal, it is not evident what is accomplished. A second suggested method is to translate the transducer relative to the fluid, or equivalently, the fluid relative to the transducer's radiating face, so that only a compressive wave is generated. Nothing is said about the fact that when the transducer is ultimately shut off, a large negative pressure pulse can be created.—WT

6,842,401

43.30.Vh SONAR BEAMFORMING SYSTEM

Alice M. Chiang and Steven R. Broadstone, assignors to TeraTech Corporation
11 January 2005 (Class 367/138); filed 19 July 2001

The sonar suite and associated electronics are described for an unmanned underwater vehicle that might be used for search, survey, mapping, or mine-field reconnaissance and hunting applications. The sonar arrays consist of a forward-looking array for obstacle avoidance, one, or preferably two, side-looking arrays, and possibly a third downward-looking array for high-resolution mapping. Suggested values for various operating parameters of the several arrays are given.—WT

6,847,588

43.30.Wi METHOD FOR CHANGING THE FREQUENCY FOR SAMPLING SONAR WAVEFRONTS

George Wallace and Paul Greene, assignors to L-3 Communications Corporation
25 January 2005 (Class 367/129); filed 16 March 2004

This patent covers a method for sampling and processing the output signals from a long line array of hydrophones so that the array can be used at frequencies above that for which the individual hydrophone spacing is one-half a wavelength, presumably without experiencing the appearance of a grating lobe in the directivity pattern.—WT

6,833,540

43.35.Sx SYSTEM FOR MEASURING A BIOLOGICAL PARAMETER BY MEANS OF PHOTOACOUSTIC INTERACTION

Hugh Alexander MacKenzie and John Matthew Lindberg, assignors to Abbott Laboratories
21 December 2004 (Class 250/214.1); filed in the United Kingdom 7 March 1997

This system measures a biological parameter, such as blood glucose, by directing laser pulses from a light guide into a soft-tissue body part, e.g., a finger, to produce a photoacoustic interaction. The resulting acoustic signal is detected by a transducer and analyzed to yield the desired information.—DRR

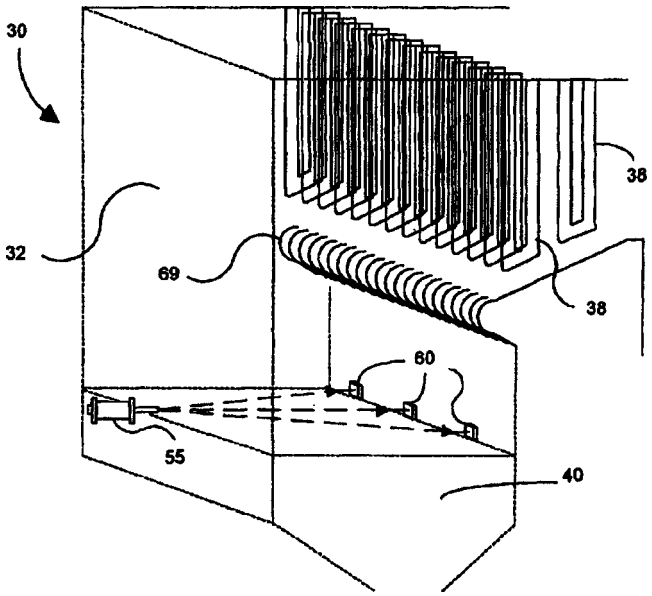
 0001-4966/2005/117(6)/3348/19/$20.00

6,834,992

43.35.Zc ACOUSTIC PYROMETER

Dean E. Draxton *et al.*, assignors to Combustion Specialists, Incorporated
28 December 2004 (Class 374/115); filed 8 September 2003

This patent is well written, succinct, and easy to follow. First, it gives us enough background information to understand the importance of accurate gas temperature monitoring in industrial coal-fired boilers. It then describes an improved acoustic pyrometer capable of determining the average gas temperature across distances of 50 feet or more. An acoustic source **55** generates a signal with high amplitude and short rise time—a brief air blast, for example. A detector adjacent to the generator responds to the onset of the



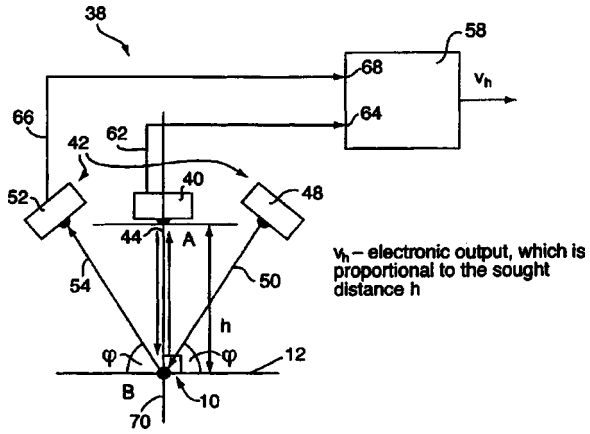
signal and one or more transducers **60** on the opposite wall of the combustion chamber receive the signal. A computerized processor utilizes digital filtering to attenuate background noise and calculate the transit time of the signal. It then calculates the temperature of the gas based on transit time. A number of additional checks are made to minimize the chance of erroneous measurements.—GLA

6,836,449

43.35.Zc ACOUSTIC METHOD AND DEVICE FOR DISTANCE MEASUREMENT

Alexander M. Raykhman *et al.*, assignors to INESA East, Limited; InESA, Incorporated
28 December 2004 (Class 367/99); filed 5 September 2003

An ultrasonic echo system is disclosed that can accurately monitor the thickness of rolled metal, paper, etc., during fabrication. In such a system, the elapsed time between the generation of an acoustic signal and the reception of its echo can be used to calculate the distance between the emitter and the target object if the speed of sound is accurately known. Unfortunately, variations in air temperature, humidity, and homogeneity can affect the



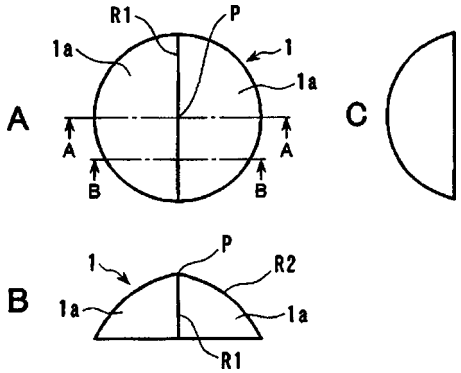
speed of sound. In this improved method, two sound paths are employed. Transducer **40** sends and receives pulses at normal incidence to target **B**. A simultaneous signal from transducer **52** is reflected at an angle back to receiver **48**. By employing a few basic trigonometric functions, sound speed can be eliminated from the calculation of distance **h**.—GLA

6,832,106

43.38.Ar ELECTROACOUSTIC TRANSDUCER

Hiroshi Sugata and Kenta Yujima, assignors to Foster Electric Company, Limited
14 December 2004 (Class 455/567); filed in Japan 21 May 2002

A hemispherical dome is used in many direct radiator transducers which are commonly called tweeters. ''This dome-shaped, hemispherical diaphragm is in axial symmetry in which the distance between the apex of the dome shape and all circumferential positions at the outer periphery along which the loudspeaker bobbin is coupled is equal all over the circumferential positions, and the vibrations transmitted from the voice coil to the outer periphery of the diaphragm is caused to concentrate at the apex in equiphase so that the resonance is apt to occur in the mode of axial symmetry so as to cause the frequency characteristic curve to involve remarkable dips specifi-



cally in the higher range of the audio frequency, whereby it has been made unable to attain excellent tone quality.'' In other words, a section of a hemispherical shell can have axial resonances. The patent describes a dome **1** that is mirror symmetric along a central arcuate ridge line **R1**, which is located along a diameter of the dome. The dome can be made from PPTA, PEN, PET, and similar films, or from titanium, aluminum, and like materials. This and a magnetic structure, that does not appear remarkable but is described in the text and listed in the claims, are said to reduce the problem quoted above.—NAS

6,836,057

43.38.Ar DRIVE MECHANISM EMPLOYING ELECTROMECHANICAL TRANSDUCER

Yoshiaki Hata, assignor to Minolta Company, Limited
28 December 2004 (Class 310/328); filed in Japan 14 September 2000

This patent describes a linear actuator based on friction-slip piezoelectric drivers that can be used for high precision applications. It does not discuss size scalability or MEMS applications.—JAH

6,837,110

43.38.Ar MICRO-MACHINED ULTRASONIC TRANSDUCER (MUT) SUBSTRATE THAT LIMITS THE LATERAL PROPAGATION OF ACOUSTIC ENERGY

David G. Miller, assignor to Koninklijke Philips Electronics, N.V.
4 January 2005 (Class 73/632); filed 30 October 2003

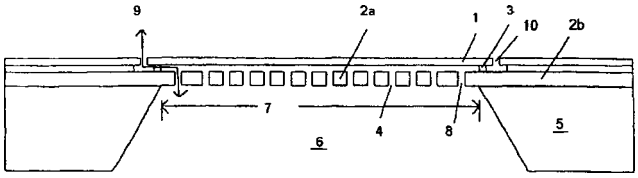
This patent teaches the art of isolating ultrasound transducers on a silicon chip by using vias (holes drilled vertically through the chip) to confine the vibrational energy. There is no technical discussion of what size or where such holes should be placed, and curious assertions are made for being able to ''dope'' the empty vias to make them conductive.—JAH

6,847,090

43.38.Ar SILICON CAPACITIVE MICROPHONE

Peter V. Loeppert, assignor to Knowles Electronics, LLC
25 January 2005 (Class 257/418); filed 8 January 2002

As stated, ''This is the Utility patent Application claims benefit of Provisional Patent Application Ser. No. 60/263,785, Filed Jan. 24, 2001.''



The patent deals with processes for manufacturing electret microphones making use of IC techniques.—JME

6,847,153

43.38.Ar POLYURETHANE ELECTROSTRICTION

Edward Balizer, assignor to The United States of America as represented by the Secretary of the Navy
25 January 2005 (Class 310/311); filed 30 May 2002

The composition and method are discussed for preparing a phase-mixed thermoplastic polyurethane that might be used as a transducer material. This material is said to have an electrostrictive strain coefficient that is two orders of magnitude greater than that of piezoceramics or piezoelectric crystals, as well as presenting an excellent impedance match to water.—WT

6,837,108

43.38.Bs INCREASING THE DYNAMIC RANGE OF A MEMS GYROSCOPE

William P. Platt, assignor to Honeywell International Incorporated
4 January 2005 (Class 73/504.16); filed 23 April 2002

This patent describes a scheme for adjusting the scale factor of a MEMS gyro by dynamically adjusting the bias voltage applied to the same pickoff electrodes used for rotation sensing. There's nothing novel about this use of capacitive feedback and sensing, as it has been in use in microphones for years.—JAH

6,841,992

43.38.Bs MEMS ENHANCED CAPACITIVE PICK-OFF AND ELECTROSTATIC REBALANCE ELECTRODE PLACEMENT

Aiwu Yue and Ronald B. Leonardson, assignors to Honeywell International, Incorporated
11 January 2005 (Class 324/162); filed 18 February 2003

This patent describes the dispersal of electrodes in a MEMS accelerometer. This is little to be learned here other than some elementary electrostatics, and the benefits of segregated electrodes for motional feedback. This is nothing new.—JAH

6,845,670

43.38.Bs SINGLE PROOF MASS, 3 AXIS MEMS TRANSDUCER

Andrew C. McNeil *et al.*, assignors to Freescale Semiconductor, Incorporated
25 January 2005 (Class 73/514.32); filed 8 July 2003

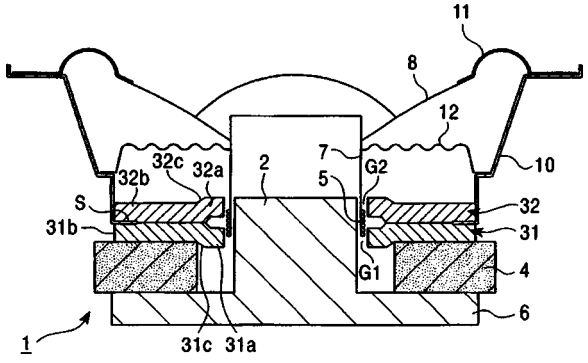
A three-axis accelerometer is described which requires fewer fabrication steps than other designs. The suspension concept is complex, and matching sensitivities in the axes is a problem in these designs. There is ample use made of polysilicon to make the connections required to the proof mass and assorted electrodes.—JAH

6,829,366

43.38.Dv MAGNETIC CIRCUIT AND LOUDSPEAKER USING THE SAME

Kei Tanabe, assignor to Alpine Electronics, Incorporated
7 December 2004 (Class 381/412); filed in Japan 21 January 2002

The linearity of this magnetic motor circuit in an electrodynamic transducer depends, in part, on how the voice coil **5** moves in the magnetic field



across the ''gap'' in which the voice coil is situated. The patent describes a motor topology that uses two top plates **31** and **32** with the inner circumference of the plates deformed so that the gap consists of two smaller axially separated subgaps **G1** and **G2**. The prior art referenced uses a machined part to produce the gap in the top plate. Among the benefits cited for this top-plate topology are that an underhung voice coil can be used and so more amplitude can be achieved for a given level of distortion. The main benefit here is that the top plate can be made from two inexpensive parts as opposed to a machined part.—NAS

6,841,922

43.38.Fx PIEZOELECTRIC RESONATOR APPARATUS WITH ACOUSTIC REFLECTOR

Robert Aigner *et al.*, assignors to Infineon Technologies AG
11 January 2005 (Class 310/335); filed 31 December 2003

This brief patent describes a technique for fabricating quarter-wave stack acoustic isolators for ZnO or AlN resonators on silicon. It does little to teach the design techniques, but merely declares success with nonstandard stack layer thicknesses. No argument is made for scaling to micron dimensions.—JAH

6,842,166

43.38.Fx PIEZOELECTRIC TRANSDUCER AND ELECTROPHORETIC INK DISPLAY APPARATUS USING PIEZOELECTRIC TRANSDUCER

Kazumasa Hasegawa and Tatsuya Shimoda, assignors to Seiko Epson Corporation
11 January 2005 (Class 345/107); filed in Japan 29 January 1999

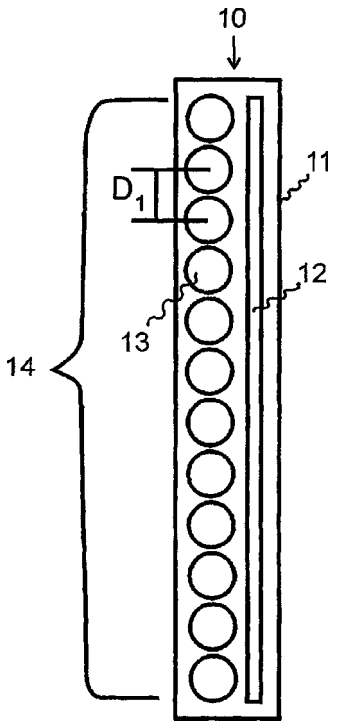
This interesting patent describes the fabrication of a resonant Rosen-type piezoresonator structure to provide high-voltage drive to individual pixels of a display that uses electrophoretic inks. The text and figures are descriptive enough that one gets the impression that it has been tried and works. Little is given in the way of design equations, however.—JAH

6,834,113

43.38.Hz LOUDSPEAKER SYSTEM

Erik Liljehag, Uppsala, Sweden *et al.*
21 December 2004 (Class 381/99); filed 3 March 2000

This patent describes a loudspeaker configuration ''which has the ability to create a homogenous sound over large distances.'' The phrase is repeated several times but never defined. Nonetheless, the basic concept is not hard to understand. For certain concert sound applications, a full-range vertical line array may be a desirable loudspeaker choice even though its vertical beamwidth varies inversely with frequency. An array of moving-coil speakers can generate plenty of low-frequency energy, but center-to-center spacing determines the usable high-frequency limit. Conversely, a ribbon



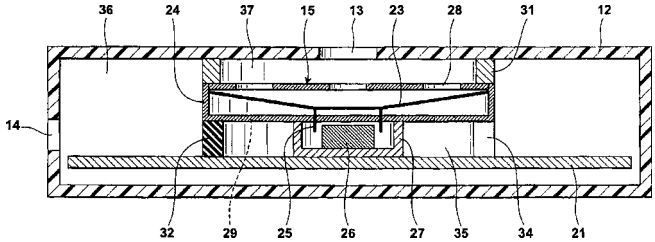
transducer operates as a true line source even at very high frequencies, but rapidly runs out of steam below 1 kHz or so. It follows that the coverage pattern of a two-way system such as that shown should be able to duplicate that of a theoretical line source at any frequency within its operating range. The configuration is not new and has been used in commercial home loudspeaker systems; however, the patent includes a detailed explanation of how to calculate the highest practical crossover frequency.—GLA

6,834,744

43.38.Ja SPEAKER SYSTEM

Nozomi Toki, assignor to NEC Corporation
28 December 2004 (Class 181/148); filed in Japan 23 April 2001

In a cellular phone, miniature speaker **15** is mounted to printed circuit board **21**. Forward radiation is emitted (or, as the patent insists, ''is output'') through chamber **37** and opening **13**. Rear radiation travels from chamber **35**



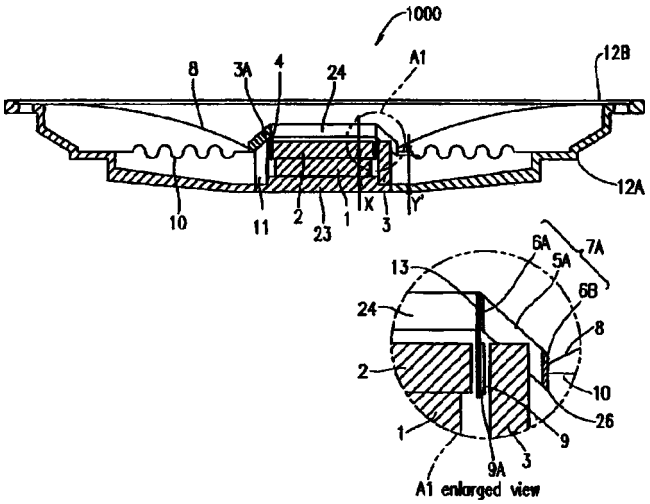
through cutout **34** to the interior of the phone housing and exits through card slot **14**. The circuitous rear path is intended to reduce interference between front and rear radiation, and thereby improve low-frequency response.—GLA

6,836,551

43.38.Ja LOUDSPEAKER

Mitsukazu Kuze *et al.*, assignors to Matsushita Electric Industrial Company, Limited
28 December 2004 (Class 381/412); filed in Japan 23 March 2000

This patent was first filed in Japan in March, 2000 and is presented as an improvement on an even earlier design. More than 40 illustrations and 62



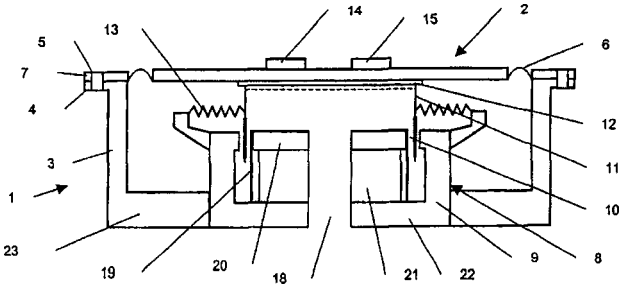
claims suggest that Matsushita rates this as something really important in the evolution of moving coil loudspeakers. To reduce overall depth, the apex of cone **8** lies well below pole pieces **2** and **3**. The motion of voice coil **9** is transferred to the cone through individual ribs **5A**, which travel up and down in slits cut into outer pole piece **3**. This would seem to be a recipe for generating some really interesting mechanical resonances, but it is described as ''a coupling member having a small mass and high mechanical strength, which can reduce the production cost associated with the mass production of various models and which can provide a sufficient vibration amplitude.''—GLA

6,839,444

43.38.Ja LOUDSPEAKERS

Christien Ellis and Nicholas Patrick Roland Hill, assignors to New Transducers Limited
4 January 2005 (Class 381/152); filed in the United Kingdom 30 November 2000

Panel-type loudspeakers operating in their bending wave range are normally designed to be driven at a single point. To meet this requirement in practice, a very small voice coil diameter is required, leading to a number of performance limitations. The 46 claims of this patent try to cover all pos-



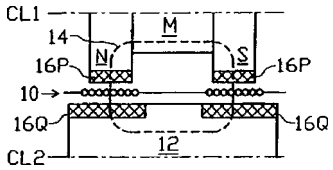
sible ways of optimizing performance using a large diameter coil. The key phrase is ''sufficient length in relation to the size of the acoustic member to represent a line drive such that the acoustic member has a mechanical impedance which has a rising trend with bending wave frequency.''—GLA

6,847,726

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

Douglas J. Button *et al.*, assignors to Harman International Industries, Incorporated
25 January 2005 (Class 381/401); filed 9 April 2003

This patent is a continuation of United States Patent 6,768,806, filed in 1999. It discloses a number of ways in which shorting rings can be intro-



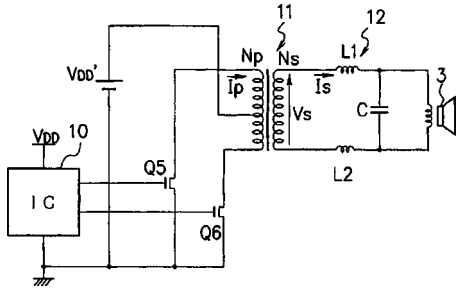
duced into the JBL dual-coil, dual-gap loudspeaker design. The patent teaches that ''a plurality of rings can be differently located so as to optimally suppress both even- and odd-order harmonic distortion and reduce voice coil inductance.''—GLA

6,844,777

43.38.Lc AUDIO AMPLIFIER

Mamoru Kitamura, assignor to Niigata Seimitsu Company, Limited
18 January 2005 (Class 330/10); filed in Japan 6 July 2001

In this class D (digital switching) audio amplifier, integrated circuit **10** is required to supply only a low-power driving signal. Additional amplification is provided by on/off transistors **Q5**, **Q6** and the voltage is further stepped up by transformer **11**. Since the transformer operates at a very high



frequency, it can be small and lightweight. The patent is written in convoluted computerish; not only are signals inputted and outputted, they are flowed.—GLA

6,740,805

43.38.Md SOUND SYSTEM AND METHOD FOR CREATING A SOUND EVENT BASED ON A MODELED SOUND FIELD

Randall B. Metcalf, Cantonment, Florida
25 May 2004 (Class 84/723); filed 30 August 2002

Most musical instruments are small, stationary sound sources. A trumpet, for example, might be modeled as a directional source producing sound waves that radiate outwardly from a single point, keeping in mind that its directional pattern may change with frequency and time. This patent teaches that such characteristics can be captured by a quasispherical array of pickup locations in the far field. Sound pressure and direction are sensed and recorded at each location during a performance (''sound event''). Instead of replacing microphones with loudspeakers to re-create the sound field, the method described here uses the recorded information to re-create the sound source. Let us assume that an array of tiny loudspeakers could be designed to reproduce the required frequency range and directionality. It seems reasonable that a finite number of such arrays could then be assembled to model a more complicated source such as a jazz combo or a choir. Numerous combinations and permutations of this approach are included in the patent, which is the latest in a series of continuations going back to United States Patent 6,239,348, filed in 1999.—GLA

6,838,056

43.38.Pf METHOD AND APPARATUS FOR SORTING BIOLOGICAL CELLS WITH A MEMS DEVICE

John Stuart Foster, assignor to Innovative Micro Technology
4 January 2005 (Class 422/100); filed 8 July 2002

This patent describes a MEMS version of a flow cytometer, a machine for sorting cells in solution. The particular application targeted here is sorting stem cells from blood, and it is done with an array of channels etched into glass, through which fluorescence is excited and observed. An array of valves does the sorting into either one of two output channels, one for stem cells and one for the rest. Electrostatically actuated valves are used in what appears to be a realizable configuration.—JAH

6,836,055

43.38.Rh PIEZOELECTRIC VIBRATOR AND FILTER USING THE SAME

Yukinori Sasaki, assignor to Matsushita Electric Industrial Company, Limited
28 December 2004 (Class 310/320); filed in Japan 11 June 2001

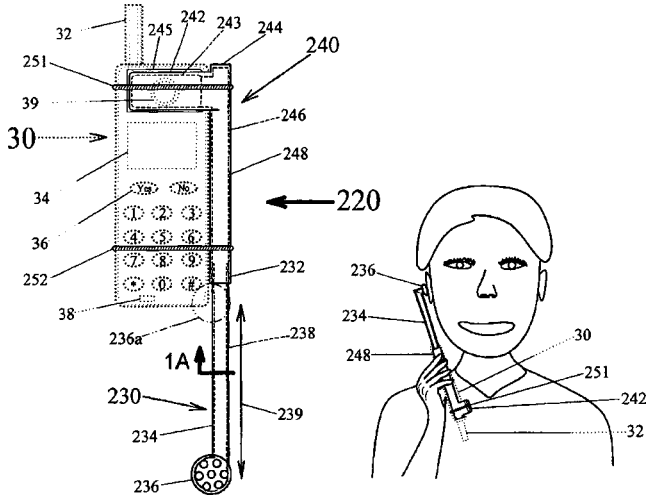
This patent discloses optimum geometries (L/W ratios) for AlN resonators that can be used as IF filters. The patent is difficult to understand as technical details have been lost in the translation.—JAH

6,825,810

43.38.Si AUDIO EXTENSION FOR WIRELESS COMMUNICATION DEVICES

Gary Dean Ragner, Gainesville, Florida *et al.*
30 November 2004 (Class 343/700 MS); filed 12 December 2002

The health problems that may be due to the electromagnetic radiation emitted by cellular phones when these devices are in use have been reported in the media. The device described in the patent allows for the use of cellular phones at a distance from the user's head. As opposed to other solutions to mitigate this problem that try to shield the unit, this patent moves the unit away from a user's head. Loudspeaker assembly **220** is



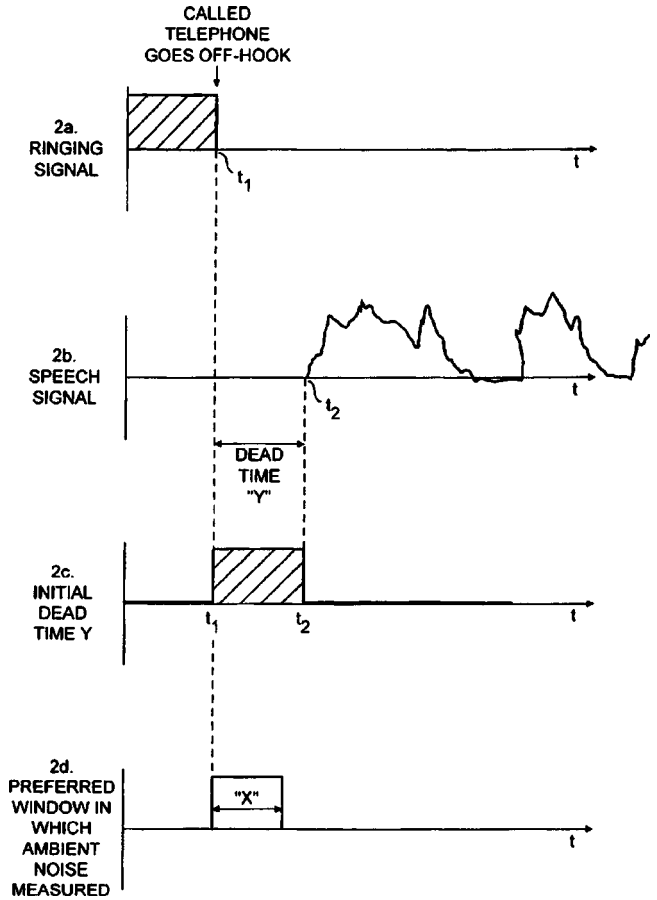
attached, or is integral, to the phone **30**, with a pickup over the cell phone's speaker feeding extension loudspeaker **239**. In use, the phone is inverted so the cell phone's microphone stays near the mouth, while the cell phone itself, and the radiation emitted from same, is removed a bit from the user's head. The patent states that among the benefits of the invention, the radiation level at the head can be reduced by an order of magnitude.—NAS

6,834,107

43.38.Si TELEPHONE APPARATUS

Jeffrey Hurst, assignor to Telefonaktiebolaget LM Ericsson
21 December 2004 (Class 379/390.01); filed in the United Kingdom 29 July 1998

The loudspeaker volume in a telephone may need to be adjusted to compensate for the ambient noise in the vicinity of the telephone. In some situations like a church, the patent states, the receiver volume may need to be turned down. At a construction site, it may need to be turned up. Some prior art methods do this by measuring the ambient noise during gaps in speech, which can be a complex process. Another method uses manual buttons. The method described in the patent adjusts the loudspeaker volume by measuring the ambient sound in a predetermined amount of time, *X*,



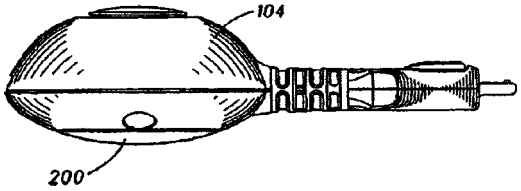
which starts when the telephone goes ''off-hook'' but is shorter than time, *Y*, which starts at the same time as *X* and ends when the called party starts to speak. A similar method is used to adjust the receiver sound level at the calling telephone. A simple flowchart shows the three steps used to accomplish this task. It is curious, though, that while this method is designed mainly for cellular and other DSP-based telephones, the term ''off-hook'' is still a valid descriptor.—NAS

6,836,676

43.38.Si SPEAKERPHONE ACCESSORY

Grant Harries Lloyd *et al.*, assignors to Motorola, Incorporated
28 December 2004 (Class 455/569.1); filed 2 November 2001

This clam-shaped cell phone accessory provides improved speakerphone capability. It is smaller than the phone itself—only about 2 inches in



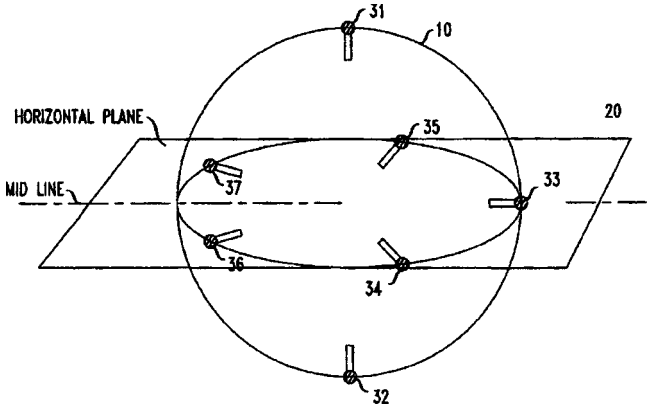
diameter—and contains a miniature loudspeaker along with associated electronics. The phone's own microphone is used during speakerphone operation.—GLA

6,845,163

43.38.Vk MICROPHONE ARRAY FOR PRESERVING SOUNDFIELD PERCEPTUAL CUES

James David Johnston and Eric R. Wagner, assignors to AT&T Corporation
18 January 2005 (Class 381/92); filed 15 November 2000

The patent describes a microphone array intended for recording direct and spatial cues in normal performance spaces. The seven microphones **31–37** lie on a sphere with a diameter of 0.9 ms of ''sound travel,'' or approximately 1 foot. The array has five microphones positioned at 72° intervals in the azimuthal plane, along with one microphone each pointing upward and downward. Although not described as such, the microphone pattern chosen appears to be roughly of hypercardioid form, at least in its front hemisphere. In playback over an equiangular, five-channel surround sound array, the contribution of the vertical-oriented microphones would be



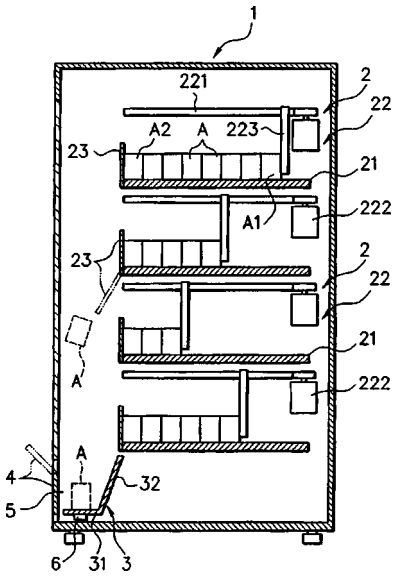
appropriately mixed into the normal five-channel array. The authors state that the microphone spacing is critical in maintaining first-arrival transient signal integrity at the central listening position, presenting to the ears a temporal pattern not unlike what might be heard at the actual recording position. Further details in the patent cover other aspects of localization and its implementation using fewer microphones. Overall, the patent is well written and is recommended reading for anyone interested in surround sound technology.—JME

6,832,695

43.40.Dx VENDING MACHINE WITH A VIBRATION SENSOR FOR CARRYING OUT GOODS

Tomomi Yamaguchi, assignor to Sanden Corporation
21 December 2004 (Class 221/195); filed in Japan 22 March 2002

This device is integrated into a vending machine that indicates by vibration whether selected merchandise is deliverable from its assigned platform. The vending machine drives a vibrating device when merchandise is selected. A vibration sensor detects the vibration of the merchandise holding



platform and a microprocessor detects whether the sensor's response is normal. If the feedback from the sensor is deemed to be normal, a delivery phase is performed. Otherwise, the delivery phase is aborted.—DRR

6,842,087

43.40.Dx THREE-TERMINAL FILTER USING AREA FLEXURAL VIBRATION MODE

Masakazu Yamauchi, assignor to Murata Manufacturing Company, Limited
11 January 2005 (Class 333/187); filed in Japan 9 November 2000

This patent describes a flexural mode filter for audio-frequency use. It is essentially a bimorph piezo element with one side operated as a generator and the other as a receiver. It resembles the old thickness-mode piezo transformer design in electroding and has potential for voltage gain. No mention is made of scaling to micron dimensions, although the center frequency would go inversely with the width of the device, making a MHz-range device feasible. Nothing really novel about this.—JAH

6,837,859

43.40.Ng SHOCK WAVE SOURCE WITH A COIL CARRIER HAVING A NON-CIRCULAR CONTOUR

Mario Bechtold *et al.*, assignors to Siemens Aktiengesellschaft
4 January 2005 (Class 601/4); filed in Germany 10 September 2001

This device generates ultrasonic shock waves for medical therapy. The patent informs us that, although existing generators operate in the range from 150 to 200 kHz, they produce unpleasant audible sound as well. A design is disclosed in which nonplanar and nonsymmetrical components inhibit resonances at audible frequencies.—GLA

6,842,088

43.40.Sk THIN FILM ACOUSTIC RESONATOR AND METHOD OF PRODUCING THE SAME

Tetsuo Yamada *et al.*, assignors to Ube Industries, Limited
11 January 2005 (Class 333/187); filed in Japan 11 May 2001

This patent describes the manufacture of a thickness-mode bulk acoustic wave resonator in the 1-GHz frequency region using ZnO material. A novel attachment technique is said to improve resonator Q and stability, but there are no data to back up the assertion. Most of the text describes the fabrication techniques.—JAH

6,842,089

43.40.Sk FILM BULK ACOUSTIC RESONATOR (FBAR) DEVICE

Joo Ho Lee, assignor to Samsung Electro-Mechanics Company, Limited
11 January 2005 (Class 333/189); filed in the Republic of Korea 21 May 2002

A thin-film bulk acoustic wave resonator is formed on silicon membranes suspended over a second layer of silicon. The air gap between the layers helps increase resonator Q and stability. The use of an air gap to reduce vibration coupling is a very old idea, and translates naturally to MEMS devices.—JAH

6,847,271

43.40.Sk COMPONENT HAVING AN ACOUSTICALLY ACTIVE MATERIAL FOR TUNING DURING OPERATION

Christian Korden *et al.*, assignors to Siemens Aktiengesellschaft
25 January 2005 (Class 333/188); filed in Germany 25 September 2000

This patent describes a scheme for obtaining tunable GHz-frequency filters in thin-film resonators by altering the dc voltage applied and operating near the Curie point of the material. The possible temperature instabilities are addressed by including a Peltier-effect temperature regulator. One wonders if all of this space and power couldn't be used more effectively for something else.—JAH

6,834,741

43.40.Tm ENGINE COVER WITH INTERNAL VIBRATION DAMPING PLATES

Alan S. Miller and Paolo Comello, assignors to General Motors Corporation; Tesma Engine Technologies
28 December 2004 (Class 180/313); filed 24 March 2003

Damping plate assemblies are mounted on the internal surfaces of an engine cover, where these assemblies are exposed to oil mist. Each assembly consists of two sheet metal plates of different thicknesses and is fastened to the engine cover by a few point supports. As the two plates bend when the assembly is subject to vibration, the plates' interfaces move relative to each other and energy is dissipated due to dry friction and oil between the interfaces.—EEU

6,834,741

43.40.Tm WIDEBAND ISOLATOR FOR ACOUSTIC TOOLS

Abbas Arian and Randall Jones, assignors to Haliburton Energy Services, Incorporated
28 December 2004 (Class 181/102); filed 7 December 2001

This patent relates to tools for logging of oil or gas wells. A transmitter in such a tool sends acoustic energy into the soil, and a receiver in the tool senses the returned acoustic signal. In order to shield the receiver from the direct acoustic signal generated by the transmitter, isolation between the two is provided. The isolation arrangement described in this patent consists in essence of a steel coil spring in a cylindrical housing, with the spaces between the coils filled with a resilient material. This arrangement is intended to be rugged enough to withstand the well logging environment.—EEU

6,836,051

43.40.Tm MOTOR

Hideharu Hiwaki *et al.*, assignors to Matsushita Electric Industrial Company, Limited
28 December 2004 (Class 310/258); filed in Japan 19 December 2002

A concentrated winding motor, such as used in refrigeration apparatus or other hermetically sealed applications, typically consists of a rotor within a stator, with the stator fitted tightly into a housing. Axial notches in the stator circumference limit transmission of vibrations from the stator to the housing, and axial holes that extend through the stator reduce the vibrations it generates.—EEU

6,840,017

43.40.Tm VIBRATION CONTROL STRUCTURE

Ikuo Shimoda and Kiyoharu Suzuki, assignors to Oiles Corporation
11 January 2005 (Class 52/167.1); filed 24 June 2002

Damping of a building's horizontal vibrations, such as those produced by wind or earthquakes, is accomplished essentially by interconnecting the girders of adjacent floors by means that dissipate energy due to relative motion between floors. The patent indicates design and connection details for means that employ dry friction or viscous damping.—EEU

6,839,694

43.40.Yq SYSTEM AND METHOD FOR WEIGHING ITEMS SUCH AS MAILPIECES IN THE PRESENCE OF EXTERNAL VIBRATION

Michael J. Kasmin and Edilberto I. Salazar, assignors to Pitney Bowes Incorporated
4 January 2005 (Class 705/407); filed 7 June 2002

The output of a load cell is fed to an analog-to-digital converter, and the resulting digital signal is low-pass filtered. A microprocessor analyzes the filtered signal to determine weights of items located on a platform that is connected to the load cell and feeds information to a postage meter. The microprocessor determines the weight of the item from the difference between the peaks and valleys of the filtered signal once vibrations have decayed so that this difference is less than a predetermined value. The processor identifies peaks and valleys from changes in the sign of the signal's derivative.—EEU

6,839,697

43.40.Yq SYSTEM AND METHOD FOR WEIGHING ITEMS SUCH AS MAILPIECES IN THE PRESENCE OF EXTERNAL VIBRATION

David J. Eaton *et al.*, assignors to Pitney Bowes Incorporated
4 January 2005 (Class 705/414); filed 28 June 2002

The basic principle of this patent is like that of United States Patent 6,839,694, reviewed above, except for the approach to calculating the weight of the item on the platform. The present patent includes an accelerometer and limit detector for generating a ''no go'' signal if the external vibrations exceed a predetermined limit. The microprocessor accepts the measured weight as valid if the vibrations do not exceed this limit and otherwise initiates an alternative weight determination method.—EEU

6,840,109

43.40.Yq VIBRATORY TRANSDUCER

Wolfgang Drahm and Alfred Rieder, assignors to Endress+Hauser Flowtec AG
11 January 2005 (Class 73/650); filed in Germany 8 May 2002

This patent describes a system for measuring the viscosity, density, and/or mass flow of fluid flowing through a pipe. The system in essence consists of a straight section of pipe that is made to vibrate in various modes. Viscosity of the fluid in the pipe is determined by subjecting the tubular test section to rotational vibrations about its axis and measuring the damping of these vibrations. Density is determined from observed resonant frequencies of flexural vibrations of the test section, and mass flow rate is ascertained from out-of-plane flexural vibrations of the test section that result from Coriolis forces generated as the section is made to vibrate in a given plane.—EEU

6,841,970

43.40.Yq DUAL-USE GENERATOR AND SHOCK ABSORBER ASSISTANT SYSTEM

Mark Zabramny, West Windsor, New Jersey
11 January 2005 (Class 320/101); filed 20 December 2002

An electromagnetic transducer is arranged in mechanical parallel with a vehicle's conventional shock absorber. A switching arrangement that is controlled by an occupant of the vehicle permits operation in two modes. In one mode, the electrical energy generated by the transducer as the vehicle bounces is used to charge a battery. In the other mode, the transducer acts in parallel with the shock absorber, increasing the effective suspension stiffness and providing a firmer ride.—EEU

6,843,128

43.40.Yq METHOD FOR DETERMINING AUTOMOTIVE BRAKE STRUCTURE VIBRATION DAMPING AND FRICTION MATERIAL BONDING

Frank Chen *et al.*, assignors to Ford Global Technologies, LLC
18 January 2005 (Class 73/574); filed 3 December 2003

A brake shoe to be tested is mounted on a resiliently supported platform and is subjected to broadband random vibration excitation. The resulting vibrations are detected by means of a laser velocimeter and subjected to frequency analysis. Damping of the brake shoe is determined from the half-power point frequencies (at which the velocity is 3 dB below that corresponding to a resonance peak).—EEU

6,843,277

43.50.Gf DUCT AND ELECTRONIC APPARATUS HAVING THE DUCT

Hiroyuki Meguro *et al.*, assignors to Seiko Epson Corporation
18 January 2005 (Class 138/125); filed in Japan 6 July 2001

This rather oddly titled patent covers a duct that is said to have a high sound-absorbing effect yet retains exhausting efficiency. The duct consists of a hollow duct body, a diverging section located in the intermediate part of the passageway, and sponges bonded to internal surfaces of the duct. The ducting and the divergent section may be as thick as 0.25 mm and are formed from sheets of paper impregnated with a synthetic resin such as polymethylpentene on its surface. It is not at all clear from the patent what type of electronic apparatus is targeted.—DRR

6,839,427

43.50.Ki METHOD AND APPARATUS FOR ECHO CANCELLER AUTOMATIC GAIN CONTROL

Mansour Tahernezhaadi and Michael J. Kirk, assignors to Motorola, Incorporated
4 January 2005 (Class 379/406.01); filed 20 December 2001

Echo cancellation is frequently employed in communication systems to combat line echo from impedance mismatches and also acoustic echo from hands-free devices. The usual methods involve computing an expected echo signal by applying a reference signal from the far-end communicator to an impulse response that is periodically determined for the transmission line. The expected echo is then subtracted from the actual echo, in the hopes that they are equal. This patent confronts a problem that often plagues the application of the above theory, *viz.*, it doesn't work very well when the input signal energy containing the echo is significantly different from the reference signal energy used to estimate the expected echo. The patent presents a detailed algorithm of low complexity for computing a gain to adaptively adjust the unwanted reference versus input power disparity internal to the echo canceller. The methods are also extended to a full-duplex scenario in which a double-talk threshold is used.—SAF

6,842,526

43.50.Ki ADAPTIVE NOISE LEVEL ESTIMATOR

Michael Walker, assignor to Alcatel
11 January 2005 (Class 381/94.1); filed in Germany 24 October 2000

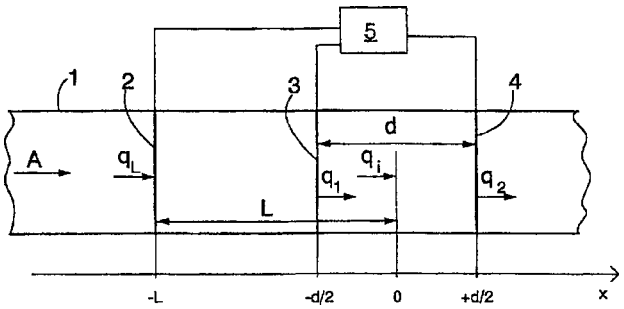
A good idea is frequently a very simple one, and this patent exemplifies the old adage. The background noise level in an acoustic signal is here taken to be the minimum in a short sequence of maximum signal excursions, so long as these excursions do not exceed a stipulated threshold (which can be set to equal a likely minimum amplitude for the desired signal). In this way, the background noise level can be adaptively found during any short pause in the desired signal, such as speech, without the need for long-term averaging and the attendant delays in adaptation. It is surprising this has not been done before, given its elegant simplicity. The method is inherently limited in its usefulness to desired signals in which some pauses are anticipated.—SAF

6,847,722

43.50.Ki METHOD AND EQUIPMENT FOR ATTENUATING SOUND IN A DUCT

Seppo Uosukainen *et al.*, assignors to VTT
25 January 2005 (Class 381/71.5); filed in Finland 12 September 1997

This is an active noise control system for attenuating noise inside a duct. Sound propagating inside a duct is detected by detector **2** and attenuated by two successive monopole elements **3** and **4** in such a manner that



both elements function as a dipole approximation. These elements are used to produce the monopole radiation needed. A dipole control signal is fed to both elements in a 180° phase shift, which effectively attenuates the noise of the flow.—DRR

6,847,723

43.50.Ki VOICE INPUT APPARATUS

Shingo Kiuchi *et al.*, assignors to Alpine Electronics, Incorporated
25 January 2005 (Class 381/94.7); filed in Japan 12 November 1998

This patent suggests that in certain kinds of noisy environments, such as a moving vehicle, the moment when a user activates a microphone is an opportune time to take account of the background noise and set the parameters of a noise-canceling scheme accordingly. A two-microphone system for keeping track of the noise and updating the appropriate cancellation filter is described; the updating of the parameters is halted when the microphone is activated. The object is to improve speech recognition accuracy in a noisy environment by canceling the noise adaptively, but it is necessary for the speech to be provided in the same noise conditions as existed when the microphone was activated. Knowledge of prior art methods in noise cancellation seems to be a prerequisite to successful replication of the technique, since numerous steps are black boxes.—SAF

6,782,095

43.60.Dh METHOD AND APPARATUS FOR PERFORMING SPECTRAL PROCESSING IN TONE DETECTION

Michael Leong *et al.*, assignors to Nortel Networks Limited
24 August 2004 (Class 379/386); filed 27 July 1999

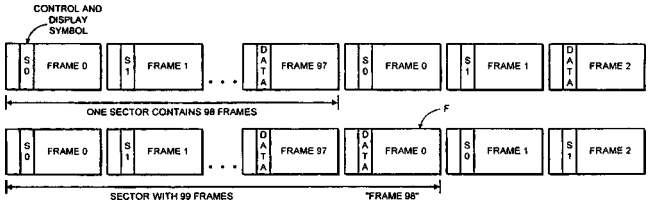
This patent describes a Fourier transform-based system for detecting and analyzing telephone signaling tones, including MF, DTMF, CPT, and international MF-R2 tones. An initial discrete Fourier transform (DFT) is performed on short subframes of the input, sampled at 8 Ksps. Magnitude information provides an initial indication that a tone may be present. If so, a quadrature signal of selected, phase-linked coefficients is passed to a second DFT, which computes a power estimate for each of the candidate signal frequencies. Phase linking the sequences of first-stage coefficients makes their sum equivalent to another transform.—DLR

6,839,312

43.60.Dh COPY PROTECTION OF DIGITAL AUDIO COMPACT DISCS

Richard A. A. Heylen and Roger Edward, assignors to Macrovision Corporation
4 January 2005 (Class 369/53.21); filed in the United Kingdom 15 February 2000

To provide copy protection, sectors on a CD are provided with a nonstandard number of frames and subcode sync patterns placed one frame later than normal. Data within each frame remain the same. These deliberately



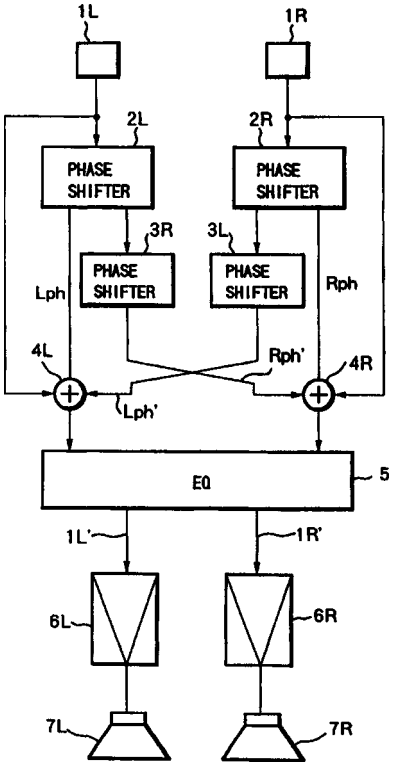
introduced errors would cause copies to have degraded sound quality via data readers but would not adversely affect audio players.—DAP

6,842,524

43.60.Jn METHOD FOR LOCALIZING SOUND IMAGE OF REPRODUCING SOUND OF AUDIO SIGNALS FOR STEREOPHONIC REPRODUCTION OUTSIDE SPEAKERS

Wataru Kobayashi, assignor to OpenHeart Limited; A Limited Responsibility Company, Research Network
11 January 2005 (Class 381/1); filed 26 May 2000

As can clearly be seen in the figure, the patent describes yet another method of introducing all-pass, variable-phase cross-talk components between a pair of stereo channels to produce a widening of the perceived



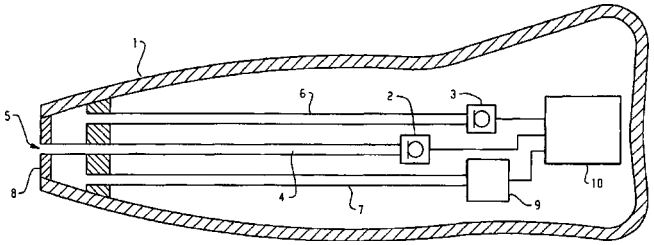
stereo stage. The approach is somewhat more complex than most, and art here lies in the fine-tuning of a number of variables.—JME

6,837,857

43.60.Qv METHOD FOR THE RECORDING OF ACOUSTIC PARAMETERS FOR THE CUSTOMIZATION OF HEARING AIDS

Alfred Stirnemann, assignor to Phonak AG
4 January 2005 (Class 600/559); filed 29 July 2002

A method is described to automatically customize hearing aid performance based on ear canal impedance measures. Two pressure microphones are enclosed in a hearing aid shell that is inserted into the ear canal for the



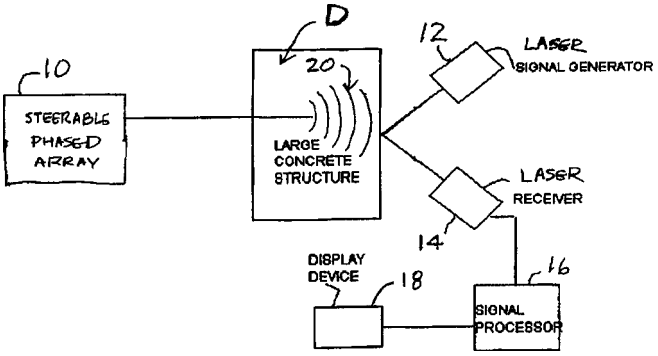
measurements. One microphone has a calibrated acoustic resistor that serves as a reference. Circuitry in the hearing aid shell analyzes the signals and calculates the impedance, distance to the eardrum, and residual volume.—DAP

6,823,737

43.60.Rw NON-CONTACT INSPECTION SYSTEM FOR LARGE CONCRETE STRUCTURES

William F. Kepler and Kurt F. von Fay, assignors to The United States of America as represented by the Secretary of the Interior
30 November 2004 (Class 73/602); filed 6 June 2001

As the patent abstract puts it, ''A phrased (sic) array acoustic source transmits an acoustic wave onto a concrete structure. A laser transmitter–receiver unit transmits a laser beam onto the surface of the structure such that the beam is modulated by the acoustic wave and reflected back to the receiver portion of the unit. The acoustic wave will be distorted by defects



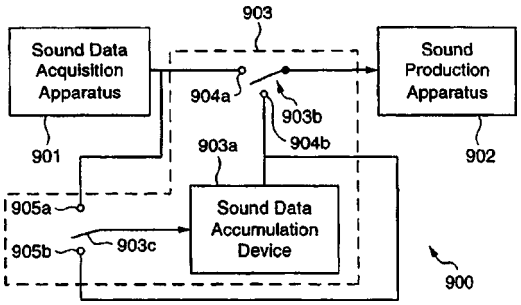
or anomalies in the concrete structure and this will affect the signal content of the received laser beam. The laser beam is ultimately converted into a two-dimensional image or a three-dimensional tomographic image for further processing.''—NAS

6,839,446

43.66.Ts HEARING AID WITH SOUND REPLAY CAPABILITY

Trevor I. Blumenau, San Francisco, California
4 January 2005 (Class 381/312); filed 28 May 2002

An add-on system is described which enables conventional analog and digital hearing aids to store at least 5 seconds of incoming sound for later playback at the same rate or at a different rate. Mode selection logic is



provided to send either current input sound data or the stored sound data to the hearing aid loudspeaker. The mode selection may be made on one or more hearing aids from a remote device via wired or wireless means.—DAP

6,839,447

43.66.Ts SYNCHRONIZED BINAURAL HEARING SYSTEM

Peter Ostergaard Nielsen and John Melanson, assignors to GN ReSound A/S
4 January 2005 (Class 381/312); filed 14 January 2003

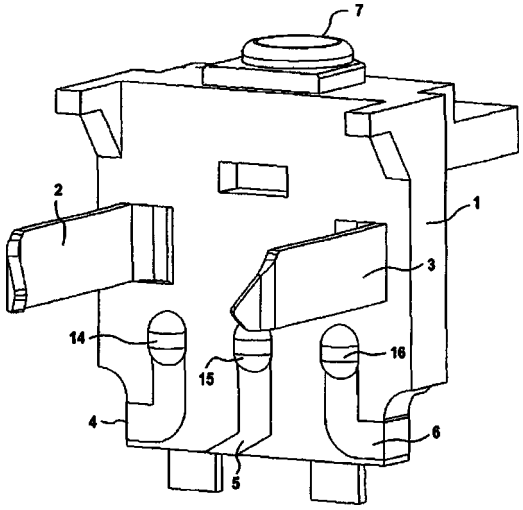
For successful implementation and control of directional beamformers implemented via bidirectional communication between two hearing aids in a binaural fitting, synchronization using a low-power, wireless, direct-sequence spread spectrum (DS-SS) technology keeps the phase error between the two hearing aids below 10 μs. The DS-SS signals may be formed from pseudorandom noise sequences.—DAP

6,843,690

43.66.Ts CONTACT DEVICE FOR HEARING AIDS

Joseph Sauer and Christian Schmitt, assignors to Siemens Audiologische Technik GmbH
18 January 2005 (Class 439/722); filed in Germany 2 April 2002

An injection-molded plastic body that houses electronic components for hearing aids contains insert-molded contacts. Wiring interconnection cir



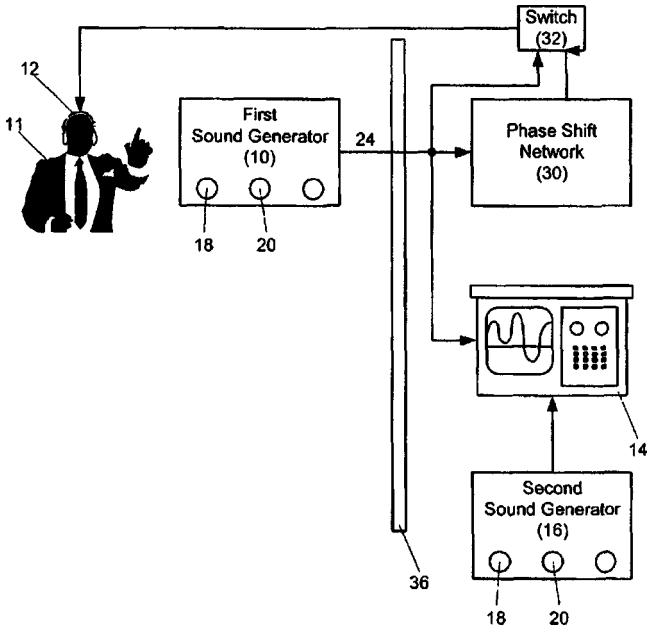
cuitry is printed onto the surface of the injection-molded plastic body.—DAP

6,846,284

43.66.Ts METHOD AND APPARATUS FOR TREATMENT OF MONO-FREQUENCY TINNITUS

Daniel S. J. Choy, assignor to Tinnitus Control, Incorporated
25 January 2005 (Class 600/28); filed 12 December 2002

Reciprocal noise cancellation of the patient's monofrequency tinnitus tone is achieved by utilizing an externally generated tone that is subjectively defined by the patient to match his/her tinnitus tone in frequency and amplitude. The externally generated sound wave is applied via earphones or a speaker system. This externally generated sound is sequentially phase



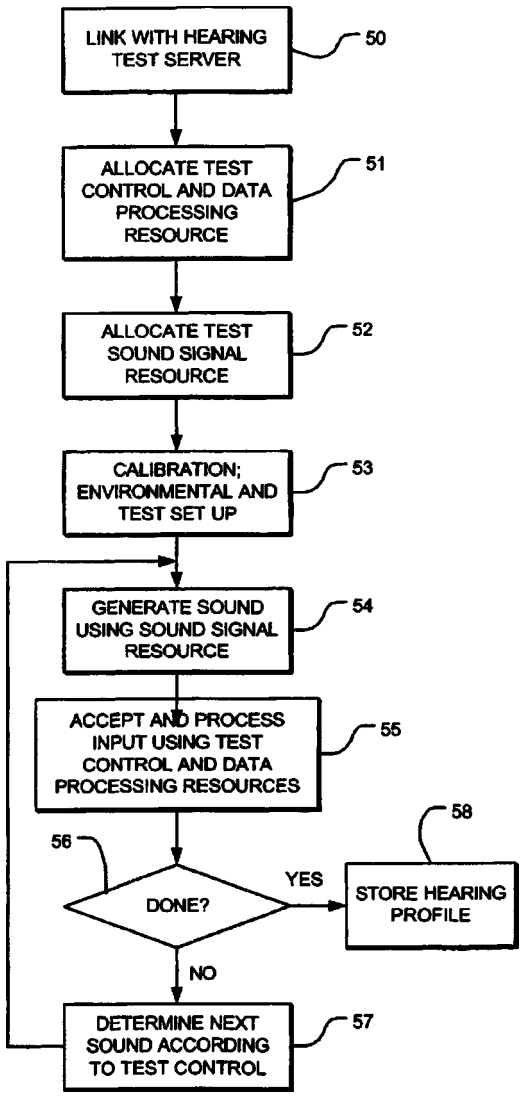
shifted through angularly shifted sequence steps in an attempt to achieve 180° phase shift of the generated signal with respect to the patient's tinnitus in order to effect noise cancellation.—DRR

6,840,908

43.66.Yw SYSTEM AND METHOD FOR REMOTELY ADMINISTERED, INTERACTIVE HEARING TESTS

Brent W. Edwards *et al.*, assignors to Sound ID
11 January 2005 (Class 600/559); filed 12 October 2001

A methodology is described for a self-administered hearing loss test via the Internet using wired or wireless communication with any of a variety of consumer electronics devices. A two-part hearing-loss computer program has a first component that is executed on the server and a second part that is executed by the remote device. After establishing a baseline threshold at the



remote device, audio stimuli to be presented during the test are selected adaptively using an *N*-alternative forced-choice interactive convergent process.—DAP

6,834,108

43.72.Dv METHOD FOR IMPROVING ACOUSTIC NOISE ATTENUATION IN HAND-FREE DEVICES

Gerhard Schmidt, assignor to Infineon Technologies AG
21 December 2004 (Class 379/406.01); filed in Germany 13 February 1998

Hands-free telephones are plagued by a problem with echo, caused chiefly by the microphone picking up the acoustic signal emanating from the loudspeaker. Traditionally, this is suppressed using a level discriminator which strongly damps one signal path at a time—thus rendering duplex communications impossible. This patent outlines a more sophisticated scheme for canceling this acoustic side-tone echo and thereby facilitating duplex telephony with hands-free devices. The first step involves applying a filter bank to the microphone signal and using the result to compute a set of estimated microphone subband signals—the description of just how is unclear. The difference between the measured and estimated microphone signals is taken as the adaptation error having two components—a baseline ''undisturbed'' error and a component due to duplex speech interference. Subband echo cancellers are then constantly adapted to the signal by prior art methods without the use of Fourier spectra, saving on computation. A

step-size variable for each subband is set in accordance with the ratio of the two mean-squared adaptation error components. Implementation of this cancellation showed that it was still incapable of completely removing the echo, and so a further Wiener postfiltering is applied which is controlled by the same step-size variable. The entire scheme, like the patent as such, is quite complicated, and it may not be sufficiently clearly described to permit replication.—SAF

6,778,960

43.72.Ja SPEECH INFORMATION PROCESSING METHOD AND APPARATUS AND STORAGE MEDIUM

Toshiaki Fukada, assignor to Canon Kabushiki Kaisha
17 August 2004 (Class 704/260); filed in Japan 31 March 2000

This patent describes a technique for computing phoneme durations for use by a speech synthesizer. In the first step, a sequence of preliminary phoneme durations is obtained by look-up from a phonetic disctionary. A prosody model computes utterance-length pitch and amplitude curves from the phonetic information. A duration adjustment step then alters the individual phoneme durations according to the preliminary durations, the computed prosodic structure, and the desired overall utterance duration. Some of the processing details are shown for a few examples in the Japanese language.—DLR

6,778,961

43.72.Ja METHOD AND SYSTEM FOR DELIVERING TEXT-TO-SPEECH IN A REAL TIME TELEPHONY ENVIRONMENT

David L. Walker *et al.*, assignors to WConect, LLC
17 August 2004 (Class 704/260); filed 16 May 2001

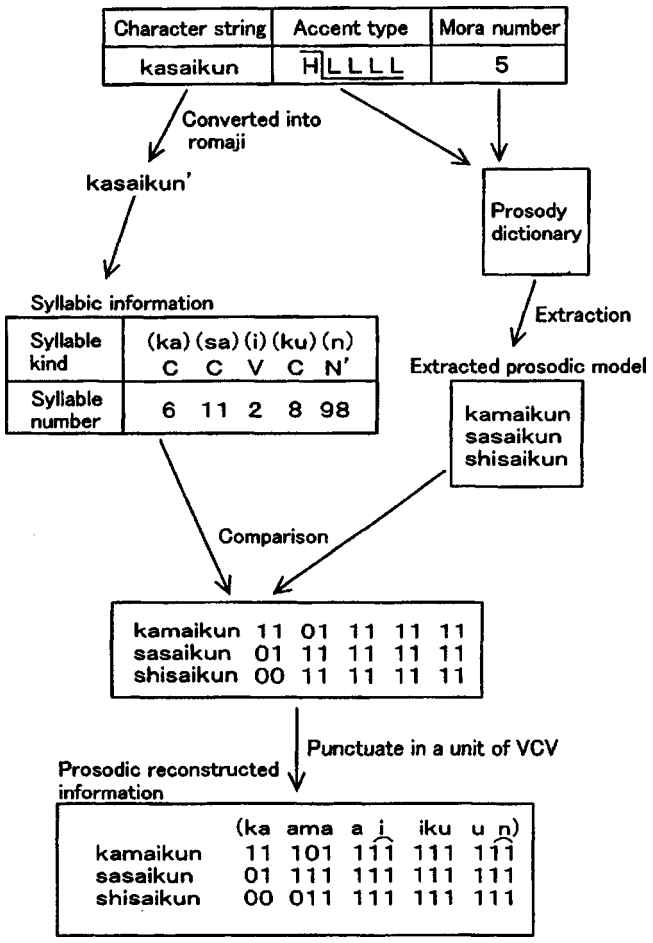
This patent basically says that multiple text-to-speech synthesis engines can be used together to speed up the synthesis of replies from an information database. To do this, the input stream of text from email, news, stock quotes, or other sources, is divided into suitable-sized chunks, which are allocated to the various processors. The chunk size would be determined, it appears, more by the required delivery rates than by any linguistic criteria. There is no mention of possible cooperation between the individual processors in the setting of various global parameters that might lend more naturalness to the overall output. More attention is given to the buffering by which the synthesized results are recombined into the final output utterance.—DLR

6,778,962

43.72.Ja SPEECH SYNTHESIS WITH PROSODIC MODEL DATA AND ACCENT TYPE

Osamu Kasai and Toshiyuki Mizoguchi, assignors to Konami Corporation; Konami Computer Entertainment Tokyo, Incorporated
17 August 2004 (Class 704/266); filed in Japan 23 July 1999

The primary object of this speech synthesis system is to be able to produce acceptably natural-sounding output when one or more of the input words (or Japanese characters) cannot be found in the phonetic dictionary. A prosody dictionary contains the character string, a mora number, an accent type, and syllable information. Candidate phoneme sequences are constructed from the available dictionary information. The accent type of each entry, either rising or falling, is used to construct a prosodic pattern. Redundancy between the syllablic structures reconstructed from the retrieved dictionary entries and the syllablic structure implied by the raw input sequence is used to fill in any gaps remaining from missing dictionary entries.—DLR



6,842,735

43.72.Lc TIME-SCALE MODIFICATION OF DATA-COMPRESSED AUDIO INFORMATION

Michele M. Covell *et al.*, assignors to Interval Research Corporation
11 January 2005 (Class 704/503); filed 13 September 2000

Audio data-compression procedures such as MP3 work by reducing the number of amplitude quantization bits, and thereby the dynamic range and amplitude accuracy, in the various frequency subbands of a filtered digitized audio signal. The reductions in quantization are performed frame by frame in accordance with perceptual models, to minimize the audibility of the reductions upon playback. This patent recognizes that, since the compression is performed with reference to the spectral domain and preserves the original frame rate, time-scale modification of the signal (such as reducing or increasing the playback rate by applying an overlap-add method) can be performed without decompressing the frames. The method itself is accordingly fairly straightforward. It is noticed that simple frequency-domain modifications, such as frequency doubling, can also be performed because the compression operations in each subband are independent.—SAF

6,778,958

43.72.Ne SYMBOL INSERTION APPARATUS AND METHOD

Masafumi Nishimura *et al.*, assignors to International Business Machines Corporation
17 August 2004 (Class 704/235); filed in Japan 30 August 1999

This speech-to-text system uses two or more so-called ''language models'' to process the results of the low-level phonetic analysis of the speech input. The models apparently consist of word spellings and word phonetic sequences, that is, typical lexical look-up information. At least one of the symbol sets used may include punctuation marks, as well as words. As processing proceeds, the recovered symbol sequences are updated such as to increase the estimated occurrence probabilities of the constructed lexical symbol sequences. The patent uses the terms ''reliability'' and ''renewal'' in ways that would usually be expressed as ''probability'' and ''update,'' respectively.—DLR

6,778,959

43.72.Ne SYSTEM AND METHOD FOR SPEECH VERIFICATION USING OUT-OF-VOCABULARY MODELS

Duanpei Wu *et al.*, assignors to Sony Corporation; Sony Electronics Incorporated
17 August 2004 (Class 704/256); filed 18 October 2000

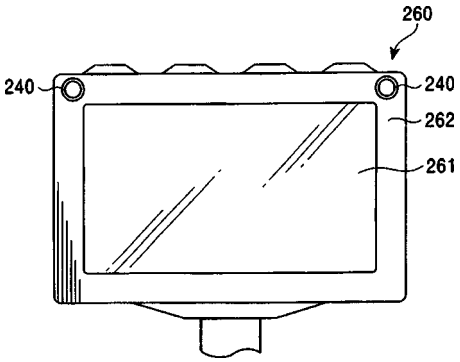
Like many current speech recognition systems, this one makes use of out-of-vocabulary models, often called garbage models. (The patent makes a minor distinction between garbage models and noise models.) Some details of the model training process described here may be novel. An initial set of garbage sound examples, including human and environment sounds, is used to train a set of HMMs. This training is iterated, producing additional models, until a desired number of models is available. The limiting criteria include the opposing effects of accuracy and processing time.—DLR

6,778,963

43.72.Ne SPEECH RECOGNITION SYSTEM FOR CAR NAVIGATION AND MICROPHONE

Toru Yamamoto *et al.*, assignors to Sony Corporation
17 August 2004 (Class 704/275); filed in Japan 19 May 2000

The emphasis in this patent is on the arrangement of one or more microphones for adequate pickup of a speech signal, while rejecting noise, in a vehicle navigation system. The solution presented is to position micro-



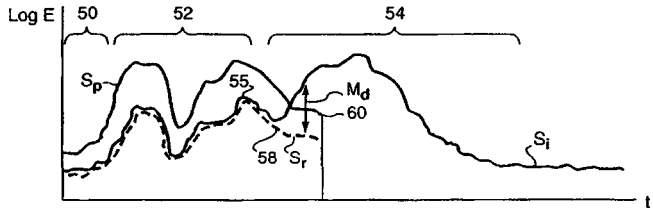
phones **240** on the edge of a display unit, at least one facing the speaker, one facing away, and possibly one on the ''other face'' of the unit. One group of claims deals with arrangements for adjusting the amount of delay in the mic signal paths, producing a beamformer pickup pattern.—DLR

6,785,365

43.72.Ne METHOD AND APPARATUS FOR FACILITATING SPEECH BARGE-IN IN CONNECTION WITH VOICE RECOGNITION SYSTEMS

John N. Nguyen, assignor to Speechworks International, Incorporated
31 August 2004 (Class 379/88.01); filed 24 July 2001

Detection of a user input during the output of a system prompt is made complicated by the possibility of line echoes. This system collects signal energy during prompt outputs, referring to these as prompt residuals. Col-



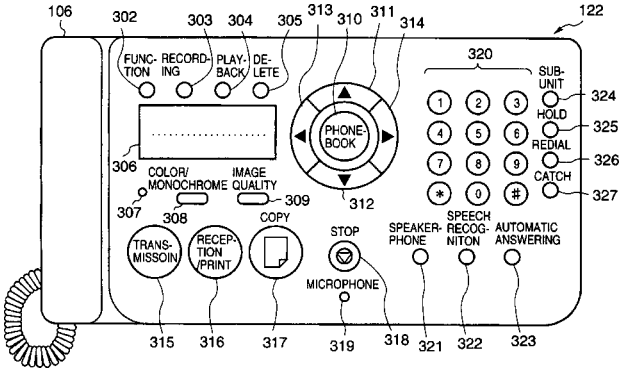
lected prompt residuals are subtracted from the input during subsequent prompts, providing a measure of whether the user might be trying to interrupt the prompt.—DLR

6,785,366

43.72.Ne APPARATUS FOR MAKING OUTGOING CALL

Hiroshi Nobuta *et al.*, assignors to Canon Kabushiki Kaisha
31 August 2004 (Class 379/88.03); filed in Japan 1 October 1998

This device recognizes speech input for the purpose of placing one or two phone calls. The patent is entirely concerned with the protocols involved in assessing the probability of correctness of the recognition results,



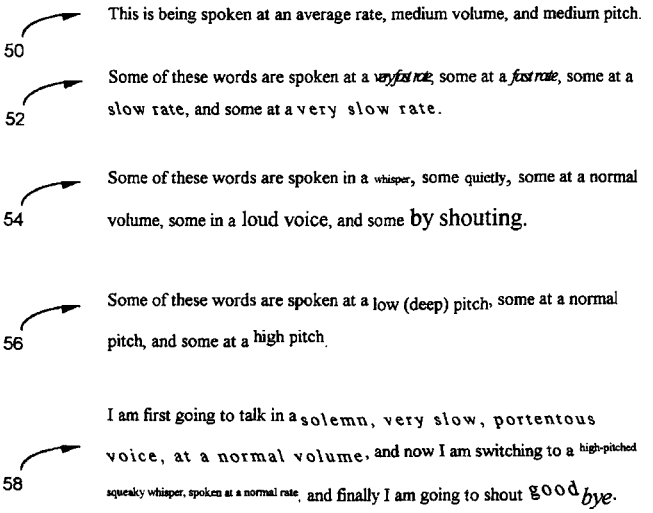
performing error recovery, and deciding whether or not the call(s) should be placed. There is extensive description of the functions performed by the various buttons on the device.—DLR

6,785,649

43.72.Ne TEXT FORMATTING FROM SPEECH

Ron Hoory and Alan Jay Wecker, assignors to International Business Machines Corporation
31 August 2004 (Class 704/235); filed 29 December 1999

This patent describes a clever way of representing various nonverbal aspects of the speech signal using standard typesetting font technology.



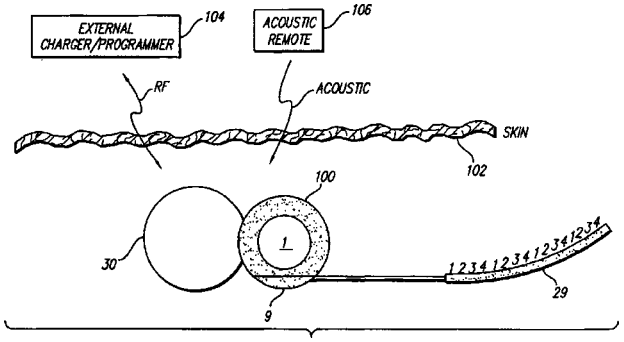
Elaborate systems of pitch and amplitude scaling and normalizing are applied to determine which of the various modes of expression should be used to display the text of a given spoken phrase.—DLR

6,826,430

43.72.Ne HIGH CONTACT COUNT, SUB-MINIATURE, FULLY IMPLANTABLE COCHLEAR PROSTHESIS

Michael A. Faltys *et al.*, assignors to Advanced Bionics Corporation
30 November 2004 (Class 607/137); filed 30 March 2001

A microphone, circuitry, and rechargeable battery for a cochlear implant are packaged in a hermetically sealed case. A detachable, programmable electrode array connects to the electronics inside the housing, allow



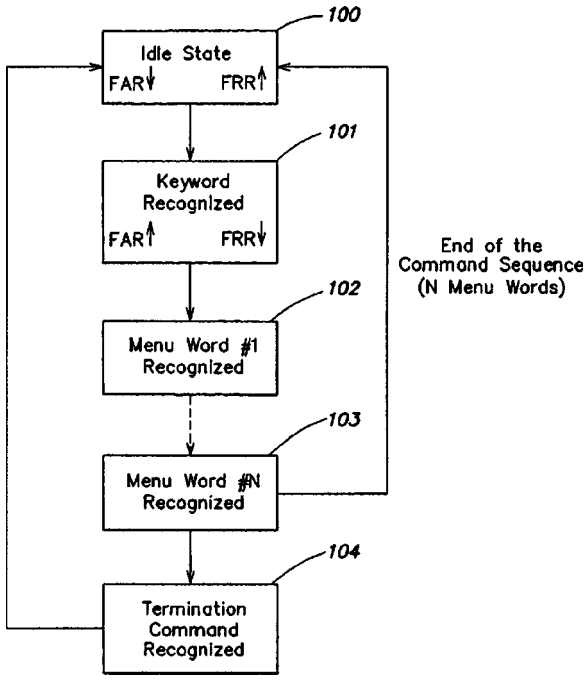
ing the completely implanted, hermetically sealed case to be surgically replaced. The number of electrodes and signal processing algorithms may be programed remotely using acoustic and/or rf control signals.—DAP

6,826,533

43.72.Ne SPEECH RECOGNITION APPARATUS AND METHOD

Bernd Burchard *et al.*, assignors to Micronas GmbH
30 November 2004 (Class 704/270); filed in Germany 30 March 2000

To improve accuracy of speech recognition, two detection thresholds are used. The first stage seeks to identify a keyword within the digitized audio signal. If a keyword has been identified, the second stage searches for a voice command. The false acceptance threshold **FAR** for voice commands



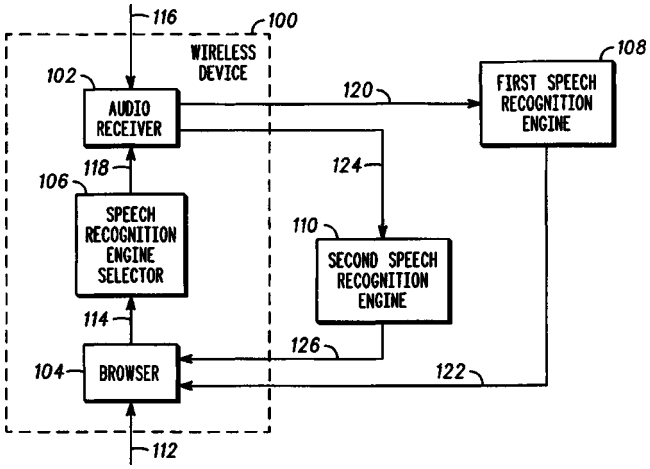
is initially lowered until a keyword is recognized. After keyword identification, the false rejection threshold **FRR** is decreased to lower the probability of an erroneous rejection of a voice command.—DAP

6,834,265

43.72.Ne METHOD AND APPARATUS FOR SELECTIVE SPEECH RECOGNITION

Senaka Balasuriya, assignor to Motorola, Incorporated
21 December 2004 (Class 704/270.1); filed 13 December 2002

An electronic device, such as a cell phone or laptop computer, contains a media type indicator that selects a first or second speech recognizer engine



depending on whether an input file is audio, video, text, or another media type. The speech recognition engines may be local or external to the electronic device.—DAP

6,839,667

43.72.Ne METHOD OF SPEECH RECOGNITION BY PRESENTING N-BEST WORD CANDIDATES

David E. Reich, assignor to International Business Machines Corporation
4 January 2005 (Class 704/240); filed 16 May 2001

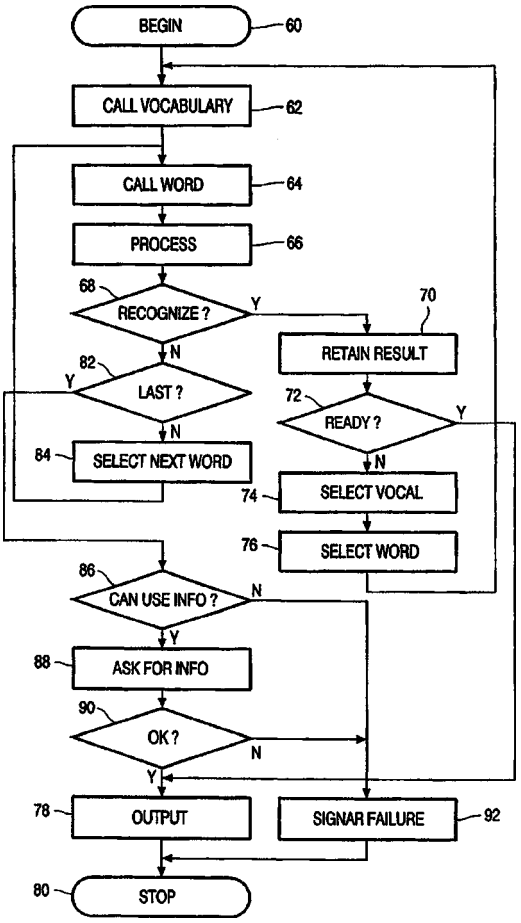
The "novel" element of this patent does not amount to much. The method described involves first attempting the correct recognition of input speech and assigning likelihood, or confidence, scores to candidate word models by entirely and admittedly prior art techniques. The patented claim is just the idea for the system to then query the user about which candidate word they actually said, by presenting choices from among those having the highest computed confidence scores. One can only marvel at such ingenuity.—SAF

6,839,668

43.72.Ne STORE SPEECH, SELECT VOCABULARY TO RECOGNIZE WORD

Jie Yung Kuo *et al.*, assignors to Koninklijke Philips Electronics N.V.
4 January 2005 (Class 704/244); filed in the European Patent Office 18 December 2000

Uninterrupted speech inputs are processed by a speech recognizer by activating instantaneously, within a relatively large overall vocabulary, only



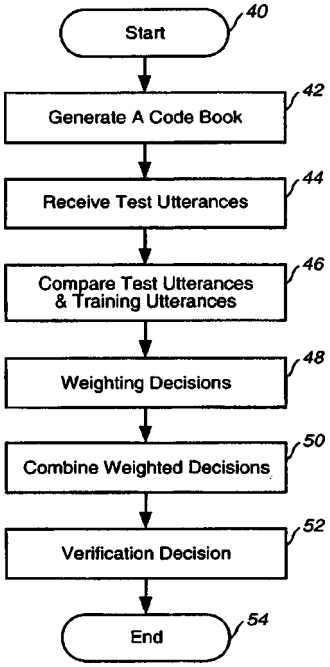
small parts that are associated with successfully recognized speech items. Vocabulary switching occurs as other speech inputs occur.—DAP

6,847,717

43.72.Ne METHOD OF ACCESSING A DIAL-UP SERVICE

Robert Wesley Bossemeyer, Jr., assignor to JBC Knowledge Ventures, L.P.
25 January 2005 (Class 379/355.01); filed 27 May 1997

A dial-up service with speech recognition is described in which digits in the first utterances of users are verified with speech verification. A code



book stores cepstral coefficients for training and impostor utterances for later comparisons.—DAP

6,837,334

43.75.Fg ACOUSTIC HORN

Dominic Le Prevost, Lower Wear, Nr Axbridge, Somerset, the United Kingdom
4 January 2005 (Class 181/177); filed in the United Kingdom 27 March 1998

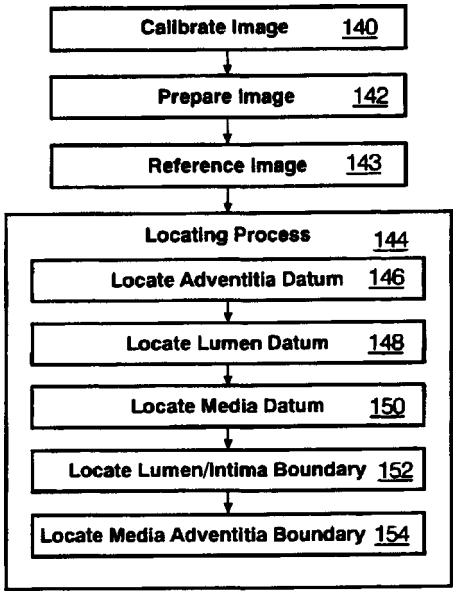
A good sheet metal worker knows how to lay out patterns for transitional couplers, i.e., a rectangular inlet to a circular outlet. This patent describes a number of methods for folding sheet material into acoustic horns that include at least one nonplanar surface. These are intended to be passed out as inexpensive party favors that provide "pleasing acoustical properties."—GLA

6,835,177

43.80.Qf ULTRASONIC BLOOD VESSEL MEASUREMENT APPARATUS AND METHOD

Terry Fritz and Helmuth Fritz, assignors to Sonosite, Incorporated
28 December 2004 (Class 600/443); filed 9 October 2003

This patent covers the methods and apparatus for processing digital images of vascular structures, in particular, the methods for interpreting ultrasonic images of the common carotid artery. In order to determine the apparent intima-media thickness (IMT), an acquired image may be referenced, establishing threshold values relating to the intensity of pixels forming images of portions of an artery wall, especially the lumen, media, and



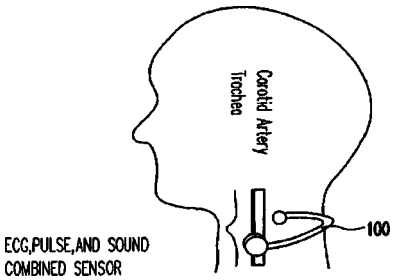
adventitia. One or more datums may be established across multiple columns of pixels bounding a portion of the image, either the lumen/intima boundary, the media/adventita boundary, or both. Within that bounded portion of the image, a procedure may search for intensity gradients indicating probable locations of the lumen/intima and media/advetitita boundaries. An IMT measurement is calculated based on the location of these boundaries and may be adjusted for sloping or tapering of the artery wall.—DRR

6,836,680

43.80.Qf DETECTOR FOR MEASURING ELECTROCARDIOGRAM, PULSE AND VOICE SOUND AT THE SAME TIME

Terry B. J. Kuo, assignor to Leadtek Research Incorporated; Terry B. J. Kuo
28 December 2004 (Class 600/513); filed 15 January 2002

The purpose of this device is to measure electrocardiogram, pulse, and voice sounds at the same time. It is designed to promote ease of use and comfort while executing a noninvasive diagnostic technique. The measured electrocardiogram signal and pulse signal are digitally processed to yield



quantitative values of the autonomic nervous system activity that regulates individual organ function and homeostasis, such as heartbeat, digestion, breathing, and blood flow. The voice sound generated from the vocal chords is not only a tool of communication, but also useful in assessing respiratory diseases; hence, it can be used as an index of noninvasive diagnosis.—DRR

6,837,855

43.80.Qf USE OF AN ULTRASONIC TRANSDUCER FOR ECHOGRAPHIC EXPLORATION OF HUMAN OR ANIMAL BODY TISSUES OR ORGANS IN PARTICULAR OF THE EYEBALL POSTERIOR SEGMENT

Michel Puech, Paris, France
4 January 2005 (Class 600/452); filed in France 18 December 1997

A high-frequency and long-focus ultrasound transducer is used to implement an echographic method of exploring human or animal tissues or organs, specifically, the macular region of the human eyeball. The ultrasonic transducer has a nominal excitation frequency greater than 20 MHz, preferably in the range of 50–80 MHz, with focal length greater than 10 mm, preferably about 25 mm, in order to render possible an echographic exploration of the posterior region of the eyeball.—DRR

6,840,913

43.80.Qf ACOUSTIC DETECTION OF GASTRIC MOTILITY DYSFUNCTION

Hansen A. Mansy and Richard H. Sandler, assignors to Biomedical Acoustic Research Corporation
11 January 2005 (Class 600/586); filed 8 March 2002

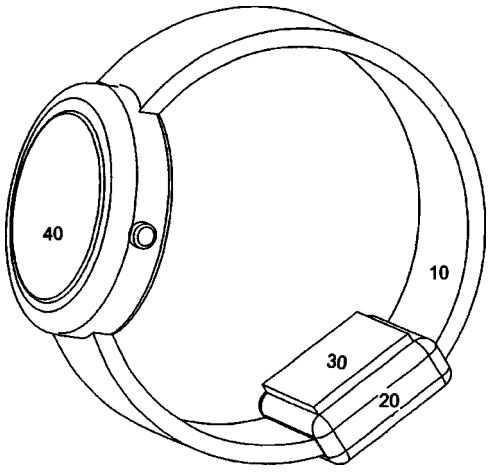
This is apparently a passive device that provides a diagnostic technique that detects, that is, ''listens to'' acoustic events at or near the surface of a patient's abdomen. The device processes electrical signals corresponding to those acoustic events to determine whether the patient is afflicted with gastric motility dysfunction. It compares the acquired acoustic information to information associated with a healthy condition.—DRR

6,843,771

43.80.Qf ULTRASONIC MONITOR FOR MEASURING HEART RATE AND BLOOD FLOW RATE

Thomas Ying-Ching Lo and Tolentino Escorcio, assignors to Salutron, Incorporated
18 January 2005 (Class 600/459); filed 15 January 2003

This ultrasonic monitor for measuring heart and pulse rates of a living subject boasts of compact size and low power consumption. It includes a module with at least one source of ultrasonic energy and a gel pad consisting of a polymer and about 50% to 95% by weight of an ultrasonic conductive



dilutant. The gel pad is positioned in direct contact between the module and the patient. An ultrasonic energy detector and associated hardware and software detect, calculate, and display a readout of the measured rate values.—DRR

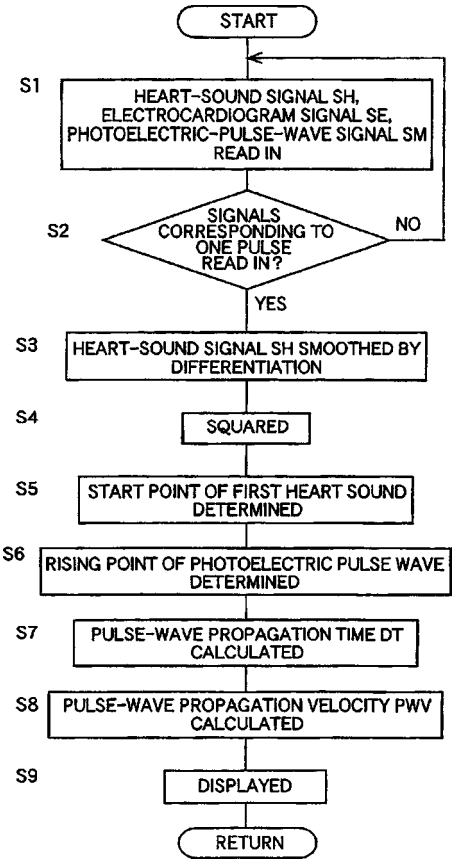
6,845,263

43.80.Qf HEART-SOUND DETECTING APPARATUS AND PULSE-WAVE-PROPAGATION-VELOCITY-RELATING-INFORMATION OBTAINING SYSTEM USING THE HEART-SOUND DETECTING APPARATUS

Keizoh Kawaguchi, assignor to Colin Medical Technology Corporation

18 January 2005 (Class 600/513); filed 2 July 2001

This heart-sound detecting apparatus has a heart-sound microphone to detect multiple sounds produced by the heart. The microphone signal wave-



form is smoothed by differentiation and amplitude squared. The start point from the first detected heart beat is established on the basis of the squared amplitude being greater than a prescribed threshold value.—DRR

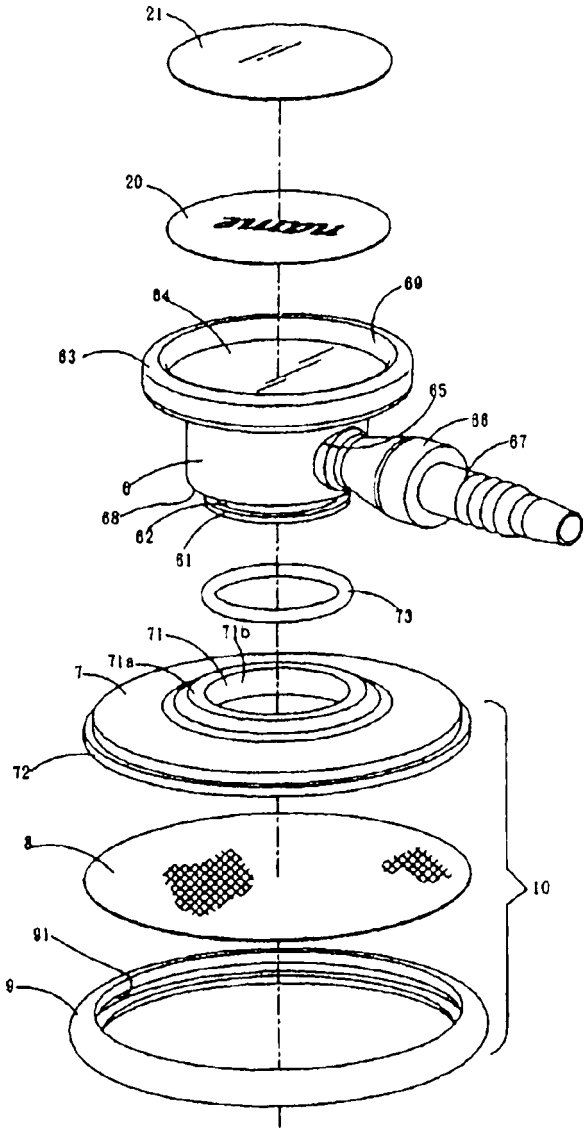
6,847,720

43.80.Qf QUICK-UPDATING STETHOSCOPE RECEIVER

Teng Ko Tseng, Taipei, Taiwan, Province of China

25 January 2005 (Class 381/67); filed 12 December 2002

The purpose of this stethoscope is to provide easy disassembly so that the listening disk (the one item in direct contact with the patient) can be



readily removed for sterilization purposes or replaced with another disk.—DRR

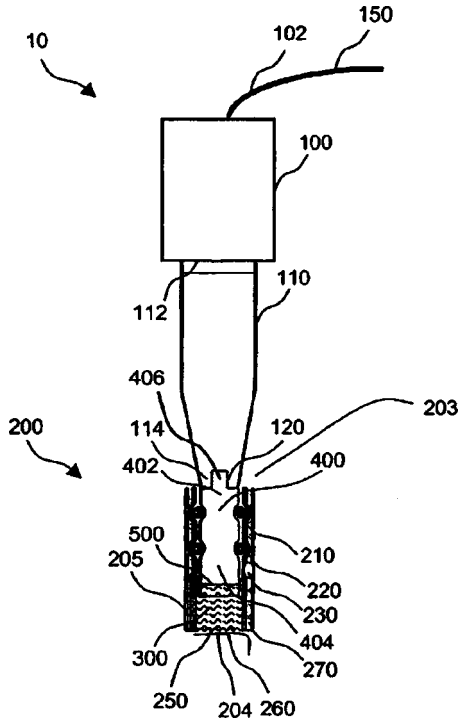
6,842,641

43.80.Sh METHOD AND APPARATUS FOR IN-VIVO TRANSDERMAL AND/OR INTRADERMAL DELIVERY OF DRUGS BY SONOPORATION

Ludwig Weimann and Richard Childs, assignors to Ultra–Sonic Technologies, L.L.C.

11 January 2005 (Class 604/20); filed 25 November 2002

This apparatus performs an *in vivo* sonoporation of skin area and transformal and/or intradermal delivery of a drug solution. A container with one end covered by a porous membrane contains the drug solution and an ultrasonic horn has a tip submerged in the solution. The horn applies ultrasound radiation to the solution with a frequency in the range of 15 kHz to 1 MHz.



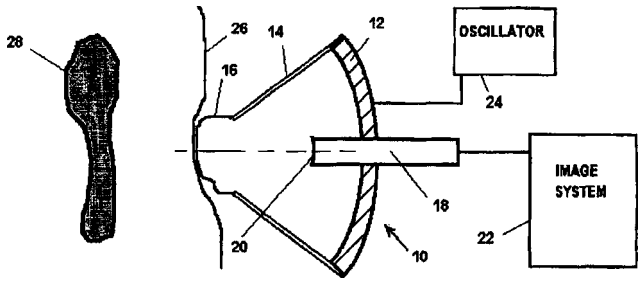
The radiation is applied at a predetermined intensity for a specific period of time and at a distance from the skin area to generate cavitation bubbles. The cavitation bubbles collapse and transfer their energy to the skin area, thus causing formation of pores in that area. The ultrasound radiation intensity generates ultrasonic jets that drive the drug solution through the porous membrane and into the formed pores in the skin area.—DRR

6,846,290

43.80.Sh ULTRASOUND METHOD AND SYSTEM

Frederic L. Lizzi *et al.*, assignors to Riverside Research Institute
25 January 2005 (Class 600/439); filed 14 May 2003

This ultrasound diagnostic system provides a means of diagnosis and treatment wherein the transducer can be located or focused according to the tissue characteristics. The system consists of a diagnostic transducer and a therapeutic transducer. The orientation of the radiation axis of each of the transducers is known with respect to the other. This assembly is positioned using the diagnostic transducer to insonify tissue in the region to be treated. Movement of the tissue with respect to the transducer assembly is detected



by the diagnostic transducer, and the focal distance of the therapy transducer is adjusted to compensate for the detected tissue movement. The tissue is then insonified using the therapy transducer with the adjusted focal distance.—DRR

6,837,853

43.80.Vj SYSTEM AND METHOD FOR USING AN ULTRASOUND TRANSDUCER WITH AN INTEGRATED TRANSDUCER INFORMATION SYSTEM

Vaughn Marian, assignor to Acuson Corporation
4 January 2005 (Class 600/437); filed 27 June 2002

The storage of ultrasound imaging system settings, transducer scanhead usage, and other information such as user comments, technical support, and transducer maintenance is described.—RCW

6,837,854

43.80.Vj METHODS AND SYSTEMS FOR USING REFERENCE IMAGES IN ACOUSTIC IMAGE PROCESSING

Thomas L. Moore and Robert Henry Barter, assignors to Barbara Ann Karmanos Cancer Institute
4 January 2005 (Class 600/437); filed 18 December 2002

Tissue and registration fiducials are insonified. Both transmitted and reflected wave information is received. A representation of the field including both the tissue and registration fiducials is derived from the received information.—RCW

6,846,289

43.80.Vj INTEGRATED X-RAY AND ULTRASOUND MEDICAL IMAGING SYSTEM

Guy M. Besson and Morgan W. Nields, assignors to Fischer Imaging Corporation
25 January 2005 (Class 600/437); filed 6 June 2003

An x-ray system and an ultrasound imaging system are integrated to permit ultrasound and x-ray imaging along essentially the same paths. In the integrated apparatus, the x-ray detector and ultrasound transducer are located on the same side of the region being imaged. The x-ray and ultrasound imaging operations may be sequential, partially overlapping, or synchronous.—RCW

LETTERS TO THE EDITOR

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

Thermodynamic cycles executed in a looped-tube thermoacoustic engine (L)

Yuki Ueda,[a] Tetsushi Biwa, and Uichiro Mizutani
Department of Crystalline Materials Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

Taichi Yazaki
Department of Physics, Aichi University of Education, Kariya, Japan

(Received 21 July 2004; revised 2 March 2005; accepted 14 March 2005)

A looped-tube thermoacoustic engine is composed of only a stack of plates, across which a steep temperature gradient is furnished. In spite of the absence of any mechanical parts like pistons and valves, the energy conversion from heat flow into acoustic power flow is performed in the stack. The thermoacoustically induced acoustic field was observed through the simultaneous measurements of pressure and velocity of the working gas, and the thermodynamic cycle executed in the stack was studied. The experimental results show that the acoustic field is tuned without any external forces so as to realize the efficient thermodynamic cycles as possible. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1904503]

PACS numbers: 43.35.Ud [RR] Pages: 3369–3372

I. INTRODUCTION

When a temperature gradient across a *stack* of plates in a looped tube exceeds some critical value, a gas column in the tube begins to oscillate spontaneously.[1–3] In this phenomenon, gas parcels in the stack undergo a thermodynamic cycle so that the energy conversion from heat flow along the stack into acoustic power flow is executed.[4–6] Hence, a looped tube with the stack having a steep temperature gradient is called ''a looped-tube thermoacoustic engine.''

The thermodynamic cycle associated with the gas parcel can be characterized by the two parameters;[6,7] one is $\omega\tau$, where ω is an angular frequency of the gas oscillation and τ is the thermal relaxation time between the gas parcel and a solid wall in the stack, and the other is a phase delay θ of the cross-sectional mean displacement $Z=\zeta e^{i(\omega t-\theta)}$ of the gas parcel relative to its pressure $P=pe^{i\omega t}$. When the gas parcel moves toward the hot end of the stack, it is naturally heated, and vice versa. As a result, the gas parcel in the stack experiences a thermodynamic cycle consisting of the compression, heating, expansion, and cooling stages. The two parameters $\omega\tau$ and θ determine the timing for the execution of these four stages. Ceperley pointed out[8] that, when the traveling wave with $\theta=\pi/2$ is formed under the condition $\omega\tau \ll \pi$, the energy conversion takes place through the thermodynamic cycle similar to the Stirling cycle, which ideally realizes the Carnot efficiency.[9,10]

We should note that, unlike a conventional engine, a looped-tube thermoacoustic engine has no mechanical parts such as pistons and valves. This means that one has no way to control the thermodynamic cycle in the thermoacoustic engine. This addresses a question as to how the thermodynamic cycle is executed in a looped-tube thermoacoustic engine. In this work, we have measured θ in the looped-tube thermoacoustic engine through simultaneous measurements of pressure and velocity[1,11–14] while changing $\omega\tau$ in the stack. We show below that θ changes without any external forces so as to make the thermodynamic cycle as efficient as possible under the condition specified by the value of $\omega\tau$. Moreover, it is shown that the pressure amplitude p and velocity amplitude u in the stack are automatically determined under a given value of $\omega\tau$. The results will be discussed in terms of the energy losses due to viscosity and thermal conduction in the stack.

II. THERMODYNAMIC CYCLE IN A THERMOACOUSTIC ENGINE

We discuss in this section how the phase delay θ and $\omega\tau$ are related to the thermodynamic cycle. The thermal relaxation time τ needed for heat exchange between a gas parcel and a solid wall is given as $r^2/(2\alpha)$, where r is the radius of

[a]Electronic mail: ueda@mizu.xtal.nagoya-u.ac.jp

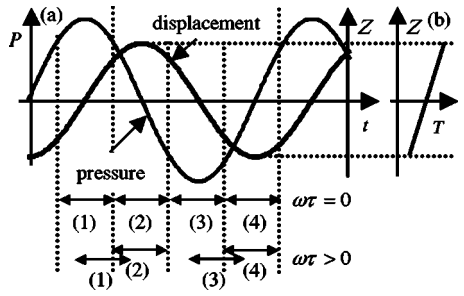


FIG. 1. (a) Time-dependence of the cross-sectional mean displacement Z and the pressure P in a gas having $\theta=\pi/2$. (b) Z dependence of temperature T in the gas. When $\omega\tau$ is zero, the gas experiences (1) heating, (2) expansion, (3) cooling, and (4) compression stages in proper timing. But, when $\omega\tau$ becomes finite, the heating and cooling stages would partially overlap the expansion and compression stages, respectively. This is because the condition $\omega\tau>0$ requires a finite time to allow heat exchange between a gas and solid wall of the stack.

the flow channel and α is the thermal diffusivity[15] of the gas. If $\omega\tau\ll\pi$, a gas parcel in the channel exchanges heat with the wall without any time delay, whereas if $\omega\tau\gg\pi$, the gas motion becomes isentropic and can no longer exchange heat with the wall. In the case of $\omega\tau\sim\pi$, the gas parcel thermally interacts with the wall with a finite time delay because of the thermal relaxation over the cross section of the flow channel.

We first consider an ideal thermodynamic cycle under the condition[8] satisfying both $\theta=\pi/2$ and $\omega\tau\ll\pi$, thereby the gas parcel being assumed to be always in thermal equilibrium with the wall of a stack. The time dependence of P and Z of a gas parcel when $\theta=\pi/2$ is illustrated in Fig. 1(a) and the Z dependence of temperature of the gas parcel in Fig. 1(b). Since $\omega\tau\ll\pi$ is fulfilled, the temperature of the gas parcel certainly coincides with that of the wall in contact. Now the gas parcel experiences following four stages: (1) the gas is heated as it moves to the hot end (heating), (2) it is depressurized in the vicinity of the hot end (expansion), (3) it is then cooled as it moves to the cold end (cooling), and (4) it is pressurized near the cold end (compression). They constitute a thermodynamic cycle similar to the Stirling one.[8,10] Therefore, the thermal efficiency under the condition satisfying $\theta=\pi/2$ and $\omega\tau\ll\pi$ ideally reaches the Carnot efficiency.[8,9]

In a real stack, however, $\omega\tau$ takes a finite value. Because of poor thermal contact between a gas parcel and a solid wall in the stack, the gas parcel would undergo heating and cooling stages with a time delay relative to its motion. Hence, suppose that θ is fixed at $\pi/2$ under a finite value of $\omega\tau$, the

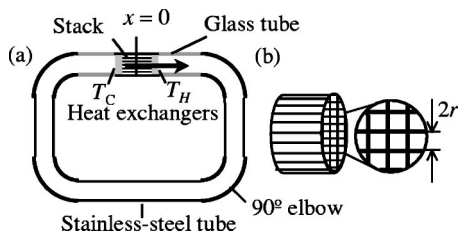


FIG. 2. (a) Schematic illustration of the thermoacoustic engine. A stack sandwiched by hot and cold heat exchangers is located in a looped tube. (b) Schematic illustration of the stack. The drawing in the circle is the overhead view of the stack.

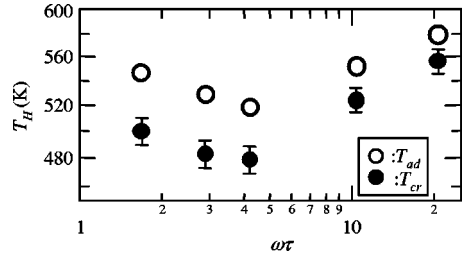


FIG. 3. Temperature at the hot end of the stack, which is plotted as a function of $\omega\tau$. The solid circle represents the critical temperature $T_{\rm cr}$, above which an acoustic wave is spontaneously induced, and the open circle the temperature $T_{\rm ad}$, at which the simultaneous measurements of the pressure and velocity were made.

start of the heating (cooling) stage would be delayed and, the expansion (compression) stage would start before the heating (cooling) stage is completely over. Thus, the heating (1) and cooling (3) stages would partially overlap expansion (2) and compression (4) stages, respectively. As a result of this improper timing over the four stages in the thermodynamic cycle, the energy conversion would become less effective compared with that through the Stirling cycle.

III. EXPERIMENTAL SETUP AND PROCEDURE

The present looped-tube thermoacoustic engine is schematically illustrated in Fig. 2(a). The looped tube is composed of Pyrex glass and stainless-steel tubes of 40 mm in inner diameter, and four 90° elbows. The total length of the looped tube, L, is 2.7 m. A ceramic honeycomb of 35 mm in length is used as a stack, which has equally spaced square pores as shown in Fig. 2(b). It is located in the looped tube and is sandwiched by two heat exchangers. One of the heat exchangers is water-cooled and kept at the room temperature T_C ($\sim$295 K), whereas the other is heated by an electric heater to the temperature T_H, which is monitored with a thermocouple. Atmospheric air is used as the working gas. The center of the stack is taken as the origin of the coordinate x as shown in Fig. 2(a). A positive direction in x is taken in the direction heading from T_C to T_H in the stack and x is normalized with respect to L.

We employed five different stacks having the edge length $2r$ of 0.6, 0.8, 1.0, 1.4, and 2.2 mm to produce a wide variety of $\omega\tau$ values in the stack. Spontaneous gas oscillation was induced with all five stacks, when T_H exceeded some critical value $T_{\rm cr}$. Independently of the choice of the stack, the frequency $\omega/(2\pi)$ of the oscillation turned out to be always about 125 Hz. By using r and ω, the values of $\omega\tau$ in the stacks are deduced to be 1.7, 2.9, 4.2, 10, and 21 at T_C.

We simultaneously measured pressure and axial velocity to determine the phase delay θ, with keeping an average pressure amplitude along the loop, $p_{\rm ave}$, to be constant at 0.68 kPa by adjusting T_H. The temperatures thus adjusted are denoted as $T_{\rm ad}$ and are plotted as a function of $\omega\tau$ in Fig. 3, where $T_{\rm cr}$ is also shown. The pressure $P=pe^{i\omega t}$ was measured with a series of pressure transducers (Toyoda Machine Works, DD-102) mounted on the tube wall via short ducts of 10 mm in length and 1 mm in inner diameter. The axial velocity was measured with a laser Doppler velocimeter along a central axis of the tube. The velocity is not uniform

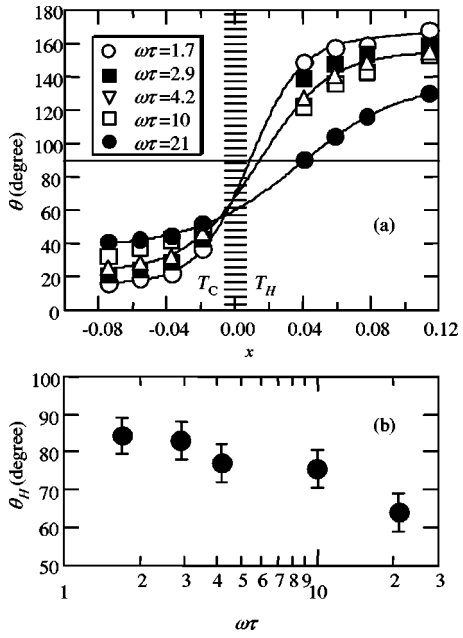


FIG. 4. (a) Distribution of the phase delay θ of the cross-sectional mean displacement Z relative to the pressure P. The position of the stack is represented by hatches. (b) $\omega\tau$ dependence of the phase delay at the hot end, θ_H, where error bars represent uncertainties of the interpolation.

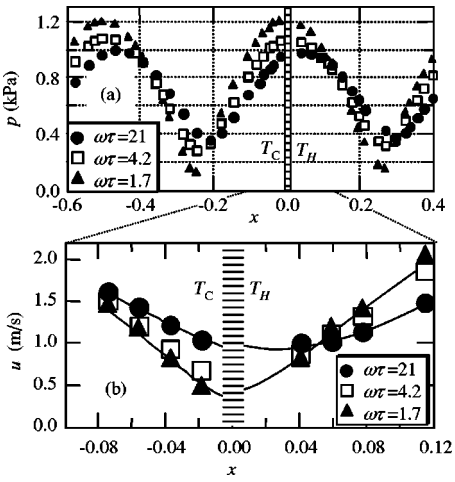


FIG. 5. Distributions of (a) the pressure amplitude p along the loop and (b) velocity amplitude u in the vicinity of the stack. Though we obtained the distributions with all five stacks, the data with the three stacks, whose $\omega\tau$ is 1.7, 4.2, and 21, are shown. The hatched area represents the position of the stack. The lines guide the eye.

along the radial direction in the tube because of viscosity of the gas. Hence, the cross-sectional mean velocity $U = ue^{i(\omega t+\Phi)}$ was calculated from the velocity measured at the center of the tube by using a laminar flow theory.[11] More details of the experimental method were described elsewhere.[11,12] The phase delay θ of the displacement Z relative to pressure P is now given by inserting the measured value of Φ into the relation $\theta = \pi/2 - \Phi$.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Phase delay of the displacement

The measured θ for the five values of $\omega\tau$ is plotted in Fig. 4(a) as a function of x over the range -0.10 to 0.12. Let us first discuss the data for $\omega\tau = 21$, which are shown by symbols of solid circle. By smoothly interpolating the data into the stack, we can obtain the phase delay $\theta_C = 55°$ and $\theta_H = 65°$ at the cold and hot ends of the stack, respectively. These phase delays are much smaller than $\pi/2$. This represents that the phase of Z in this case is advanced in comparison with that of Z in the case of $\theta = \pi/2$. It is important to note that in the stack with $\omega\tau = 21$, the heating and cooling stages of the thermodynamic cycle are delayed as mentioned in Sec. II. The reduced value of θ relative to $\pi/2$ we obtained compensates for the time delay in the heating and cooling stages. As its consequence, in the stack with $\omega\tau = 21$, the four stages, the heating, expansion, cooling, and compression stages, can be executed with the timing similar to that in the ideal case of $\omega\tau \ll \pi$ and $\theta = \pi/2$ shown in Fig. 1. However, it should be noted that the Stirling cycle can no longer be realized in this case because of the irreversible heat exchange between the gas parcel and the wall of the stack.

We now consider the other four cases, where smaller values of $\omega\tau$ are employed. As can be seen in Fig. 4(a), θ_C is not measurably dependent on the value of $\omega\tau$. Instead, θ_H strongly depends on $\omega\tau$. In Fig. 4(b), θ_H is plotted as a function of $\omega\tau$. As $\omega\tau$ is decreased, i.e., the time delay in the heating and cooling stages is reduced, θ_H in Fig. 4(b) monotonically increases and approaches 90°. In response to the decrease of the time delay, a value of θ needed for the execution of the four stages with the same timing as that in the Stirling cycle should approach 90°. Therefore, the experimental results can be taken as a demonstration that θ is naturally adjusted so as to perform the four stages with the best timing for a given value of $\omega\tau$. It is important to note that this adjustment of θ is made without any external tuning operation of θ by means of any moving parts. Based on the experimental results and discussion, we conclude that in a looped-tube thermoacoustic engine, θ is automatically adjusted to realize as efficient thermodynamic cycle as possible under a given condition of $\omega\tau$.

B. Pressure and velocity amplitudes

Figure 5(a) shows the x dependence of the pressure amplitude p measured under the condition of $p_{\mathrm{ave}} = 0.68$ kPa for the choice of different values of $\omega\tau$. It was found that the acoustic field in the looped tube is obviously characteristic of the standing wave field having pressure nodes and antinodes, though θ in the stack approaches the traveling wave phase $\theta = \pi/2$ with decreasing $\omega\tau$. In this subsection, we will discuss the $\omega\tau$ dependence of the acoustic field in terms of the energy losses caused by viscosity and thermal conduction of gas in the stack.

As $\omega\tau$ decreases and hence, the flow channel in the stack is obviously narrowed, the energy loss due to viscosity becomes serious. The viscous loss is known to be proportional to the square of the velocity amplitude u.[6] Here, we focus on the $\omega\tau$ dependence of u in the stack. In Fig. 5(b), the measured distributions of u in the vicinity of the stack are presented. We can see that u at the stack is decreased with decreasing $\omega\tau$. The decrease of u in the stack contributes to

the suppression of the increase of the viscous loss due to the decrease of $\omega\tau$.

Next, we consider the energy loss due to the thermal conduction of gas. This energy loss takes place, when the thermal relaxation due to imperfect thermal contact between a gas parcel and a wall occurs. Hence, in contrast with the viscous loss, the thermal conduction loss vanishes under the condition of $\omega\tau \ll \pi$. Since the loss is proportional to p^2,[5,6] we focus on p at the stack, where gas parcels irreversibly exchange heat with a wall. As can be seen in Fig. 5(a), p at the stack decreases with increasing $\omega\tau$. This fact means that when $\omega\tau$ is increased, the thermal conduction loss is automatically suppressed by the reduced pressure amplitude in the stack.

Summarizing this subsection, we reach the conclusion that u and p in the thermoacoustic engine are adjusted so as to minimize energy losses due to both viscosity and thermal conduction in the stack for a given value of $\omega\tau$.

V. CONCLUSIONS

We have constructed the looped-tube thermoacoustic engine by adopting the five different stacks characterized by different values of $\omega\tau$. In order to see the thermodynamic cycle in these stacks, we have measured the phase delay θ of the cross-sectional mean displacement Z relative to the pressure P. It is found that, when $\omega\tau$ is decreased, θ in the stack approaches $\pi/2$ without external tuning and that, when $\omega\tau$ is increased, θ is decreased. This is taken as evidence that the thermoacoustic engine is automatically tuned to realize the efficient thermodynamic cycle as possible under a given condition of $\omega\tau$. In addition, we revealed that u and p are also adjusted to minimize the energy loss due to viscosity and thermal conduction in the stack.

ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research No. 14005659 (2004).

[1] T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, "Traveling wave thermoacoustic engine in a looped tube," Phys. Rev. Lett. **81**, 3128–3131 (1998).

[2] S. Job, V. Gusev, P. Lotton, and M. Bruneau, "Acoustic streaming measurements in annular thermoacoustic engines," J. Acoust. Soc. Am. **113**, 1892 (2003).

[3] C. M. de Blok, "Thermoacoustic system," Dutch Patent, International Application Number PCT/NL98/00515 (1998).

[4] J. C. Wheatley, T. Hofler, G. W. Swift, and A. Migliore, "An intrinsically irreversible thermoacoustic heat engine," J. Acoust. Soc. Am. **74**, 153–170 (1983).

[5] G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1988).

[6] A. Tominaga, "Thermodynamic aspects of thermoacoustic theory," Cryogenics **35**, 427–440 (1995).

[7] T. Biwa, U. Mizutani, Y. Tashiro, M. Kotsuka, and T. Yazaki, "Experimental demonstration of thermoacoustic energy conversion in a resonator," Phys. Rev. E **69**, 066304 (2004).

[8] P. H. Ceperley, "A pistonless Stirling engine—traveling wave heat engine," J. Acoust. Soc. Am. **66**, 1508–1513 (1979).

[9] P. H. Ceperley, "Gain and efficiency of a traveling wave heat engine," J. Acoust. Soc. Am. **72**, 1688–1694 (1982); "Gain and efficiency of a short traveling wave heat engine" *ibid.* **77**, 1239–1244 (1985).

[10] S. Backhaus and G. W. Swift, "A thermoacoustic Stirling heat engine," Nature (London) **399**, 335–338 (1999).

[11] T. Yazaki and A. Tominaga, "Measurement of sound generation in thermoacoustic oscillations," Proc. R. Soc. London, Ser. A **454**, 2113–2122 (1998).

[12] T. Biwa, Y. Ueda, T. Yazaki, and U. Mizutani, "Work flow measurements in a thermoacoustic engine," Cryogenics **41**, 305–310 (2001).

[13] T. Biwa, Y. Ueda, T. Yazaki, and U. Mizutani, "Thermodynamical mode selection rule observed in thermoacoustic oscillations," Europhys. Lett. **60**, 363–368 (2002).

[14] Y. Ueda, T. Biwa, T. Yazaki, and U. Mizutani, "Acoustic field in a thermoacoustic Stirling engine having a looped tube and resonator," Appl. Phys. Lett. **81**, 5252–5254 (2002).

[15] The thermal diffusivity of gas is defined as $\alpha = \kappa/c_p\rho_m$, where κ, c_p, and ρ_m are thermal conductivity, isobaric heat capacity per unit mass, and mean density of the gas, respectively.

Beamforming with air-coupled surface waves around a sphere and circular cylinder[a] (L)

Gilles A. Daigle,[b] Michael R. Stinson, and James G. Ryan[c]
Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

(Received 9 November 2004; revised 22 February 2005; accepted 7 March 2005)

Beamforming using a microphone array mounted on the surface of a rigid or impedance sphere and cylinder is discussed. The introduction of a structure into the space occupied by or immediately adjacent to the array gives the array an effectively larger aperture and the possibility of enhanced performance. Additional improvements can be obtained through the introduction of a surface that permits air-coupled surface waves to propagate. Since surface waves travel at a speed less than the speed of sound in free space, the aperture of the array can be effectively further increased.
[DOI: 10.1121/1.1898598]

PACS numbers: 43.38.Hz, 43.20.Fn, 43.60.Fg [AJZ] Pages: 3373–3376

I. INTRODUCTION

A circular array steered to targets around its periphery is a popular geometry when the array main beam must be steered over a full 360° azimuth. A conference telephone application is a prime example. This choice of geometry avoids front–back ambiguity for talkers located on opposite sides of the array and allows a fairly uniform main beam to be steered over a full angle of 360°. For example, Meyer[1] has recently discussed the concept of phase modes to generate a desired beam pattern in the frequency range from 300 to 5000 Hz for a circular microphone array mounted around a rigid sphere.

In this paper, we first discuss the acoustical mechanisms associated with mounting a microphone array around a structure when using a simple beamforming algorithm. It is shown that mounting a circular microphone array around a rigid sphere or cylinder, through diffraction effects and the increased travel time[2] for the acoustical signals to circumnavigate the diffraction structure, gives the array an effectively larger aperture. The larger aperture can modify the characteristics of the beam pattern, such as beamwidth, sidelobe level, and directivity.

The paper then discusses the introduction of a surface treatment that permits air-coupled surface waves to propagate around the structure. Surface waves over an impedance plane are expected when the imaginary component of the surface impedance is a *springlike* reactance and is greater than the resistive component.[3,4] Surface waves can travel above the surface at a speed much less that the speed of sound in free space, depending on frequency and surface impedance. Reference 5, for example, describes how passive directivity can be achieved from air-coupled surface waves, where it is shown how directional receivers can be designed by generating surface waves above a lattice of cavities restricted to a strip of finite width embedded in an otherwise rigid surface. Raspet *et al.*[6] discuss the penetration of surface waves into shadow zones. Thus, a further modification to the characteristics of the beam pattern can be expected from the effectively large aperture provided by the reduced sound speed of the surface waves.

II. SCATTERING BY SIMPLE SHAPES

Simple shapes such as spheres and cylinders are chosen for convenience in the analysis. Exact analytical solutions are available to describe the acoustic diffraction, allowing predictions of beamforming behavior. These choices are not meant to represent an optimum design, but rather as a vehicle to illustrate concepts.

A. Scattering by an impedance sphere

An exact analytical solution for the sound field from a point source scattered by a rigid sphere is given in Bowman *et al.*[7] It is straightforward to extend the solution to the case of an impedance sphere.[8,9]

Consider a sphere of radius a with specific surface impedance $Z/\rho c = 1/\beta$, where β is the normalized specific admittance (see Fig. 1). There are N microphones equally spaced on the equator of the sphere. In the case of a point source located at (r_0, θ_0, ϕ_0) the total sound field around the sphere is, for $r < r_0$

$$\Phi = i \sum_{n=0}^{\infty} (2n+1) P_n(\cos\theta_s) h_n^{(1)}(kr_0) \times [j_n(kr) - R_n h_n^{(1)}(kr)], \quad (1)$$

where $\cos(\theta_s) = \sin(\theta_0)\cos(\phi - \phi_0)$, k is the wave number, j_n and h_n are spherical Bessel functions, P_n are Legendre functions, and

$$R_n = \frac{j_n'(ka) + i\beta j_n(ka)}{h_n'(ka) + i\beta h_n(ka)}. \quad (2)$$

In Eq. (2), the prime indicates derivatives. When $\beta \to 0$, Eq. (1) reduces to the expression found in Ref. 7 in the case of a hard sphere.

[a] Portions of this work were presented in "Air-coupled surface waves and Beamforming microphone array mounted on a sphere and circular cylinder," Proceedings of the 18th ICA, Kyoto, Japan.

[b] Electronic mail: gilles.daigle@nrc-cnrc-gc.ca
[c] Now with Gennum Corporation, 232 Herzberg Road, Suite 202, Ottawa, ON K2K 2A1, Canada.

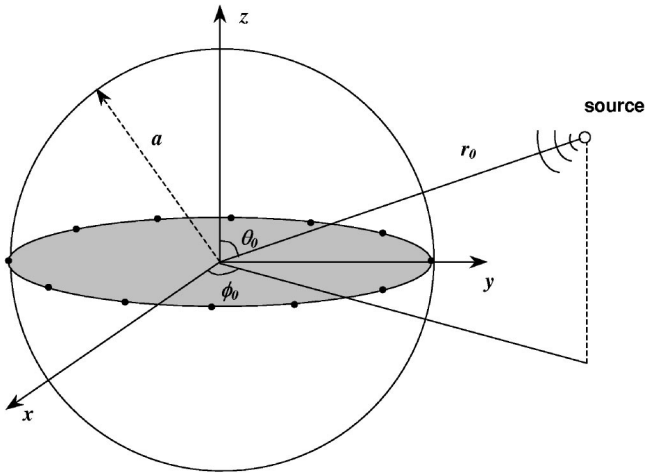


FIG. 1. Array of microphones mounted around the equator of a sphere.

B. Scattering by an impedance cylinder

Consider an infinite cylinder of radius $\rho=a$ centered along the z axis (see Fig. 2, where we have defined circular cylindrical coordinates such that $x=\rho\cos\phi$, $y=\rho\sin\phi$, $z=z$). In the case of a point source the exact solution for the sound field around a hard cylinder contains the infinite sum of a complex integral expression,[7] making computations difficult. Computations of sound fields and beam patterns using the exact solution and a far-field approximation show that essentially the same results are obtained once the source is at least a few wavelengths from the cylinder. Thus, it is straightforward to show that in the case of plane-wave incidence, the total field around an impedance cylinder is

$$\Phi=\sum_{n=0}^{\infty}\cos(n\phi)\varepsilon_n(-i)^n[J_n(k\rho\sin\theta) - R_nH_n^{(1)}(k\rho\sin\theta)], \tag{3}$$

where

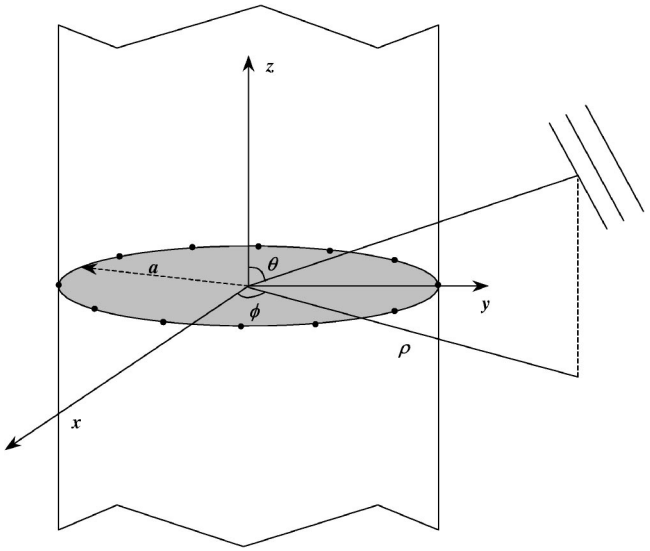


FIG. 2. Array of microphones mounted around a cylinder.

$$\varepsilon_n=\begin{cases}1, & n=0\\ 2, & n\geqslant 1,\end{cases} \tag{4}$$

and

$$R_n=\frac{J_n'(ka\sin\theta)+i\beta J_n(ka\sin\theta)}{H_n^{(1)\prime}(ka\sin\theta)+i\beta H_n^{(1)}(ka\sin\theta)}. \tag{5}$$

In Eqs. (3) and (5), J_n and $H_n^{(1)}$ are Bessel functions and the prime indicate derivatives and β is again the normalized specific admittance. When $\beta\to 0$, Eq. (3) reduces to the solution for a hard cylinder.

III. CALCULATIONS

A. Beamforming and directivity index

Assume N number of equally spaced microphones mounted in a circular array, on the equator of a sphere or around the circumference of a cylinder. A beam pattern can be calculated by multiplying the normalized sound field, G_m, observed at each microphone position m with an appropriate beam-steering coefficient, W_m, and then summing the contributions of each microphone

$$B(\theta,\phi)=\frac{1}{N}\sum_{m=1}^{N} W_mG_m. \tag{6}$$

Beam-steering coefficients can be obtained by any one of a number of array-processing strategies. In the case of a circular array in free field, the coefficients W_m in Eq. (6) will be chosen to implement simple delay-sum beamforming for the purpose of the present paper. In the case of the circular array mounted on the sphere or cylinder, the most straightforward option is to choose coefficients W_m that correspond to a delay-sum beamformer—i.e., coefficients that have unit amplitude and a phase that compensates for the time of flight (including the effects of diffraction) at each microphone. If G_{m0} is the sound field at microphone position m when the source is in the steering direction (θ_0,ϕ_0) of the array, then

$$W_m=G_{m0}^*/|G_{m0}|, \tag{7}$$

where the* indicates the complex conjugate. No attempt is made here to find optimized beam-steering coefficients since the main purpose is simply to illustrate concepts. The sound field at each microphone position m is obtained by setting $r=a$ in Eq. (1) in the case of a sphere, or $\rho=a$ in Eq. (3) in the case of a cylinder and appropriate angles θ, φ.

The directivity is the ratio of the beam-pattern response in the steering direction to the average response over all directions. The Directivity Index is 10 times the logarithm of directivity

$$\mathrm{DI}=10\log_{10}\frac{|B(\theta_0,\phi_0)|^2}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}|B(\theta,\phi)|^2\sin\theta\,d\theta\,d\phi}. \tag{8}$$

B. Surface impedance of the sphere or cylinder

The conditions for the existence of air-coupled surface waves are discussed in detail in Ref. 3. In short, surface

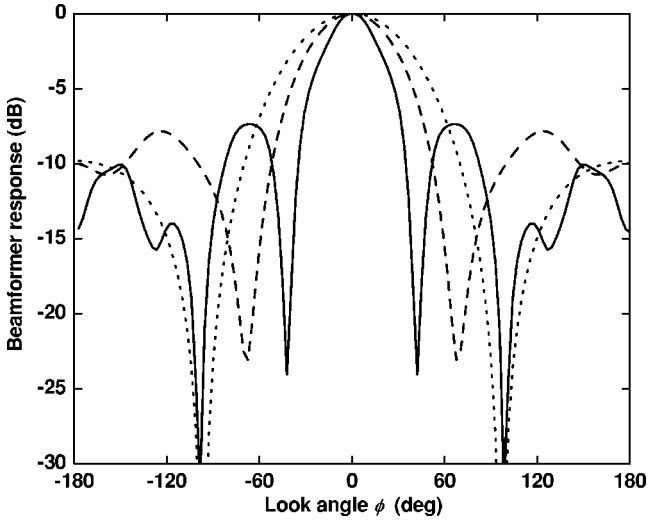


FIG. 3. Beam pattern at 660 Hz when microphones are mounted on a sphere. $N=8$, $a=13.3$ cm. Impedance sphere (solid) with $\mathrm{Re}(Z)=0.1$ and $\ell=10$ cm, hard sphere (dashed), free field (dotted).

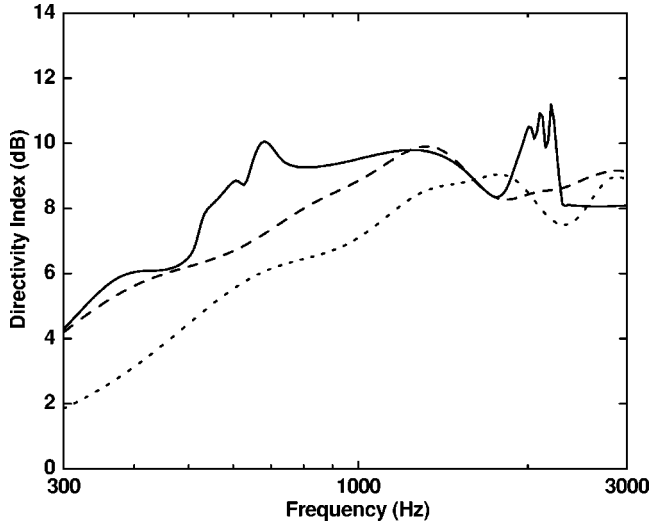


FIG. 4. Directivity Index when microphones are mounted on a sphere. $N=8$, $a=13.3$ cm. Impedance sphere (solid) with $\mathrm{Re}(Z)=0.1$ and $\ell=10$ cm, hard sphere (dashed), free field (dotted).

waves are expected when $\mathrm{Re}(Z)<\mathrm{Im}(Z)$ and $\mathrm{Im}(Z)>0$, assuming a time convention $\exp(-i\omega t)$. That is, the surface must have a ''springlike'' reactance. For the purpose here, in order to illustrate concepts, an idealized specific impedance is assumed

$$Z/\rho c=\mathrm{Re}(Z)+i\cot(k\ell). \tag{9}$$

In practice, Eq. (9) approximates the impedance of a lattice of cavities.[3] In Eq. (9), ℓ is the depth of the cavities, k is the wave number, and ρc is the characteristic impedance of air. In the extreme case when $\mathrm{Re}(Z)=0$, strong standing waves are produced around the sphere or cylinder with large peaks and dips in amplitude. This in turns makes it difficult to control the beam characteristics over a broad frequency range. If $\mathrm{Re}(Z)$ is too large, the sound field and beam characteristics collapse to the case of the hard object. By choosing the appropriate value for $\mathrm{Re}(Z)$, the standing waves are dampened and it will be shown that improved beam characteristics are possible.

IV. RESULTS AND DISCUSSION

We assume an array of eight microphones equally spaced around the equator of a sphere. The radius of the sphere is $a=13.3$ cm. A point source is located at ($r_0=3m$, $\theta_0=\pi/2$, $\phi_0=0$). The beam patterns for a frequency of 660 Hz are shown in Fig. 3. The look angle is the azimuth angle, ϕ, and the beam is steered toward the source. The dotted curve is the response of the circular array in free space calculated from simple delay-sum beamforming. The dashed curve is the response when the eight microphones are mounted equally spaced around the equator of a hard sphere. The solid curve is the response obtained by setting $\mathrm{Re}(Z)=0.1$ and $\ell=10$ cm in Eq. (9). The small resistive component of the impedance is chosen to represent realistic values (see Ref. 3; in practice, there is a weak frequency dependence). The increased travel time due to diffraction around the hard sphere produces a narrower beam, while the reduced speed of the surface wave produces an even narrower beam pattern.

The Directivity Index is shown in Fig. 4 in the case of the sphere. The dotted curve is simple delay-sum beamforming when the array is in free space. The dashed curve is the result in the case of the hard sphere. With the exception of frequencies around 1800 Hz, there is an increase in the Directivity Index at all frequencies. The Directivity Index in the case of the impedance sphere is shown by the solid curve. Surface waves are generated at frequencies when $\cot(k\ell)>0$. This occurs over two regions in Fig. 4 (at frequencies below about 900 Hz and again between about 1800 and 2500 Hz) and an increase in the Directivity Index is observed. When $\cot(k\ell)<0$, no surface waves are generated and the Directivity Index is comparable to that of a rigid sphere.

Changes to the radius of the sphere and/or the imped-

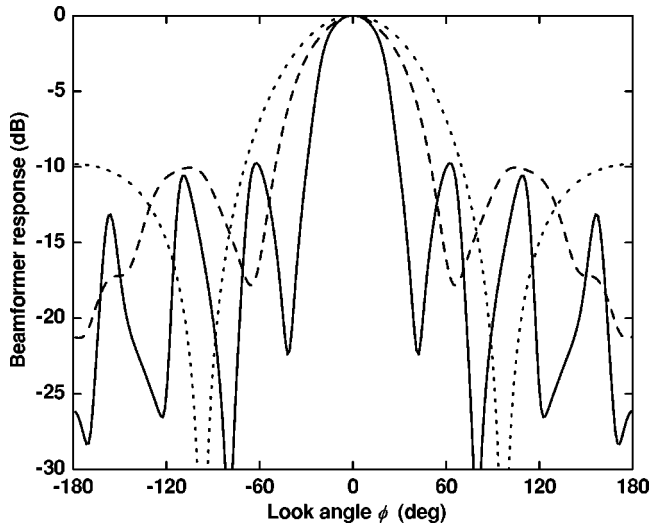


FIG. 5. Beam pattern at 660 Hz when microphones are mounted on a cylinder. $N=8$, $a=13.3$ cm. Impedance cylinder (solid) with $\mathrm{Re}(Z)=0.1$ and $\ell=10$ cm, hard cylinder (dashed), free field (dotted).

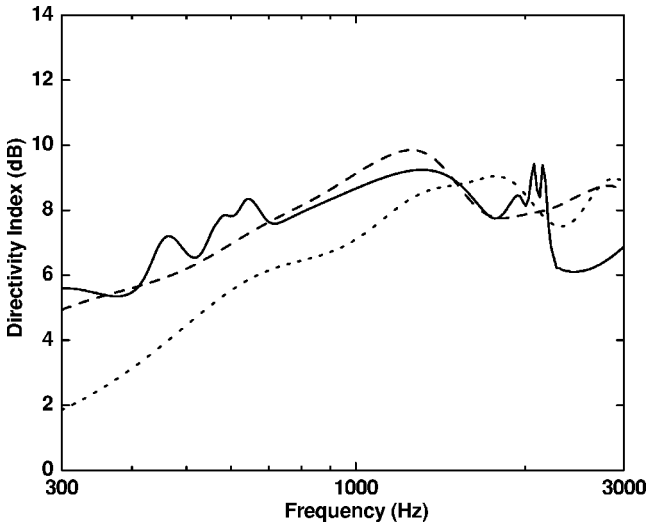


FIG. 6. Directivity Index when microphones are mounted on a cylinder. $N=8$, $a=13.3$ cm. Impedance cylinder (solid) with $\mathrm{Re}(Z)=0.1$ and $\ell=10$ cm, hard cylinder (dashed), free field (dotted).

ance produce considerable variation in the observed Directivity Index. The particular radius and impedance were used here to obtain the combined effect of the sphere and surface wave to produce an increase in the Directivity Index over the narrow-band telephone frequencies between 300 to approximately 3000 Hz.

The curves in Fig. 5 are the corresponding results when the eight microphones are mounted around the circumference of a cylinder of the same radius. Again, the beam pattern progressively narrows going from free field (dotted curve), to hard cylinder (dashed curve), to impedance cylinder (solid curve) with $\mathrm{Re}(Z)=0.1$ and $\ell=10$ cm.

The corresponding Directivity Index in the case of the cylinder is shown in Fig. 6 in the case of free field (dotted curve), the hard cylinder (dashed curve), and the impedance cylinder (solid curve). Overall, the increase in Directivity Index in the case of the hard cylinder is comparable to results found in the case of the sphere. The impedance cylinder is able to provide further increase in Directivity Index at some frequencies. Any differences in the Directivity Index between the sphere and cylinder can be attributed to the different behavior of the objects for source positions in the $y-z$ plane [see Eq. (8)].

V. CONCLUSION

Mounting an array on a physical structure increases the travel time to each microphone, giving the array an effectively larger aperture. If the surface of the structure is treated to generate air-coupled surface waves, this can effectively increase the aperture further. A hard sphere and a hard cylinder provide comparable improvement in beamwidth and Directivity Index. Both the impedance sphere and impedance cylinder were able to provide further enhancement. In the case of the chosen geometry in this paper, the combined effect of the sphere and surface waves narrows the beam pattern and increases the Directivity Index between 300 and 3000 Hz, by as much as 5 dB at some frequencies in the case of the sphere.

ACKNOWLEDGMENT

The authors would like to thank David I. Havelock for helpful discussions.

[1] J. Meyer, "Beamforming for a circular microphone array mounted on spherically shaped objects," J. Acoust. Soc. Am. **109**, 185–193 (2001).
[2] The diffraction and scattering of sound by a hard object, including the concept of creeping waves having slightly retarded phase velocity, is discussed in detail in A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 9.
[3] G. A. Daigle, M. R. Stinson, and D. I. Havelock, "Experiments on surface waves over a model impedance plane using acoustical pulses," J. Acoust. Soc. Am. **99**, 1993–2005 (1996).
[4] M. R. Stinson and G. A. Daigle, "Surface wave formation at an impedance discontinuity," J. Acoust. Soc. Am. **102**, 3269–3275 (1997).
[5] W. Zhu, G. A. Daigle, and M. R. Stinson, "Experimental and numerical study of air-coupled surface waves generated above strips of finite impedance," J. Acoust. Soc. Am. **114**, 1243–1253 (2003).
[6] R. Raspet and G. E. Baird, "The relationship between upward refraction above a complex impedance plane and the spherical wave evaluation for a homogeneous atmosphere," J. Acoust. Soc. Am. **89**, 107–114 (1991).
[7] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).
[8] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[9] D. S. Jones, *Acoustic and Electromagnetic Waves* (Oxford University Press, Oxford, 1986).

Active control of energy density in a one-dimensional waveguide: A cautionary note (L)

Ben S. Cazzolato,[a] Dick Petersen,[b] Carl Q. Howard,[c] and Anthony C. Zander[d]
School of Mechanical Engineering, The University of Adelaide, SA, 5005, Australia

(Received 24 September 2004; revised 10 February 2005; accepted 29 March 2005)

Acoustic energy density has been shown to be a highly effective cost function for active noise control systems. Many researchers have used the sound field in a one-dimensional waveguide to trial their control strategies before moving onto more realistic three-dimensional sound fields. This letter aims to shed some light on the observations made in the early papers on one-dimensional energy density control and also shows that some of the analysis was incorrect and the conclusions reached may be flawed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1920213]

PACS numbers: 43.50.Ki [KAC] Pages: 3377–3380

I. INTRODUCTION

The popularity of energy density sensors for use in active noise control systems has grown strongly in recent years, with investigations on both one-dimensional systems[1–9] and three-dimensional systems.[10–19] Acoustic energy density, defined as the sum of the acoustic potential and kinetic energy densities at a point, is commonly estimated using the two-microphone method[1] and is given by (see the Appendix)

$$E_D(z)=\frac{p^2(z)}{2\rho c^2}+\frac{\rho v^2(z)}{2}, \tag{1}$$

where z is the location, ρ is the density of the fluid, and c is the speed of sound in the fluid. The mean of the outputs from the two microphones provides an estimate of the acoustic pressure p, the square of which, when normalized by $1/2\rho c^2$, is equal to the acoustic potential energy density at a point. The pressure difference between the two microphones is used to calculate the acoustic particle velocity v, the square of which, when normalized by $\rho/2$, is equal to the acoustic kinetic energy density at a point.

Researchers investigating active noise control have often used the sound fields inside one-dimensional waveguides to investigate various control strategies prior to moving to realistic three-dimensional sound fields. The research into active noise control using energy density as a cost function is no exception. Reviews of the current literature show two main energy density control strategies in one-dimensional reactive fields: global control using a single control source[1–4] and local control using two control sources.[5–8] For all these cases the one-dimensional sound field has been modeled using a rigid-walled one-dimensional modal model. It will be shown that this modeling technique has obscured the true results and often misled researchers into believing their results to be accurate.

The work presented here has used a traveling wave solution as an alternative formulation. This solution can be shown to coincide to the modal solution for rigid-walled rectangular cavities.[20] Solving this algebraically, as opposed to past research which used numerical solutions of the modal formulation, it will be seen that all research to date has been severely compromised by numerical noise. The added advantage of the traveling wave model is the computational efficiency in which it may be used for any arbitrary termination condition, unlike the modal solution, which may require a very large number of modes for nonrigid terminations.

II. MODELING THE ACOUSTIC RESPONSE OF A RIGID-WALLED FINITE DUCT

Consider the hard-walled one-dimensional waveguide of finite length L and arbitrary termination conditions described by Φ_1 and Φ_2 as shown in Fig. 1. The pressure response in the duct at a point z arising from a source with unit volume velocity located at z_s is given by the summation of all possible direct and reflected wave components[21,22]

$$\begin{aligned}p(z,z_s)=\frac{\rho cT}{2S}[&e^{-jk|z_s-z|}+e^{-jk(z_s+z)}e^{-2\Phi_1}\\&+e^{jk(z_s+z)}e^{-2jkL-2\Phi_2}\\&+e^{jk|z_s-z|}e^{-2jkL-2\Phi_1-2\Phi_2}],\end{aligned} \tag{2}$$

where k is the acoustic wave number, c is the speed of sound, ρ is the density of air, S is the duct cross section, assumed to be small relative to the wavelength such that only plane waves propagate, and T is the modal reverberation factor given by

$$T=\frac{1}{1-e^{-2jkL}e^{-2\Phi_1}e^{-2\Phi_2}}, \tag{3}$$

where the termination phasors $\Phi_{1,2}=\pi\alpha_{1,2}-j\pi\beta_{1,2}$ (Ref. 23) are related to the reflection coefficients by $R_{1,2}=e^{-2\Phi_{1,2}}$,[23] where the exponential form results in a compact expression for Eqs. (2) and (3). The appropriate values for α and β yield specific termination conditions. For example, a rigid, totally reflective termination is achieved by setting $\alpha=\beta=0$ ($\Phi=0$), while an anechoic termination is achieved by making the real part of Φ infinite, namely $\alpha=\infty$.

[a]Electronic mail: benjamin.cazzolato@adelaide.edu.au
[b]Electronic mail: cornelis.petersen@adelaide.edu.au
[c]Electronic mail: carl.howard@adelaide.edu.au
[d]Electronic mail: anthony.zander@adelaide.edu.au

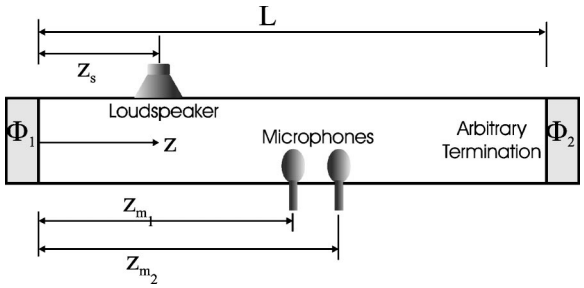


FIG. 1. Schematic of the 1D system under investigation. The duct is length L with arbitrary termination conditions described by the complex phasors Φ_1 and Φ_2. There is a volume velocity source at a distance z_s from end Φ_1 and two microphones at z_{m_1} and z_{m_2}.

Looking at Eq. (2), it can be seen that the sound field is made up of two traveling waves: a forward-traveling wave (first two terms) and backward-traveling wave (second two terms). The infinite number of reflections from the two end caps is accounted for by the modal reverberation factor in Eq. (3).

It should be noted that it has been assumed that the sound field is entirely one-dimensional, and that evanescent propagation of cross modes has not been included. Consequently, there exist no near-field effects adjacent to sources.

III. ENERGY DENSITY CONTROL

The traveling wave formulation will now be applied to two scenarios modeled previously: global control with one control source and local control using two control sources. Like previous work, causality has been neglected when formulating the control law. This is acceptable when modeling a feedforward control strategy such as the FX-LMS algorithm. The optimal strength of the control source(s), $\mathbf{q}_s$, is found by minimizing the following Hermitian quadratic form cost function:[24]

$$J=\mathbf{q}_s^{\mathrm{H}}\mathbf{A}\mathbf{q}_s+\mathbf{q}_s^{\mathrm{H}}\mathbf{b}+\mathbf{b}^{\mathrm{H}}\mathbf{q}_s+c, \tag{4}$$

where the Hermitian control matrix $\mathbf{A}=\mathbf{Z}^{\mathrm{H}}\mathbf{Z}$, the vector $\mathbf{b}=\mathbf{Z}^{\mathrm{H}}\mathbf{p}_p$, and the scalar $c=\mathbf{p}_p^{\mathrm{H}}\mathbf{p}_p$. The latter is the (weighted[17]) sum of the squared pressures at the error sensor locations for the primary field, $\mathbf{p}_p$. This weighting is such that energy density is minimized. The individual terms of the control source transfer function matrix $\mathbf{Z}$ are given by Eq. (2). Provided that the control matrix $\mathbf{A}$ is positive definite and full rank, the optimal control source strengths and corresponding minimum value of the cost function are given by[24]

$$\mathbf{q}_{s,\mathrm{opt}}=\mathbf{A}^{-1}\mathbf{b}, \quad J_{\mathrm{opt}}=c-\mathbf{b}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{b}. \tag{5}$$

A. One control source

Consider the case of a single control source investigated in Refs. 1–4, where the energy density was estimated using the two-microphone technique[1] (which is briefly summarized in the Appendix). Using simulations based on a modal model, previous researchers found that when the source was located upstream of the two microphones ($z_s<z_m$) the sound-pressure levels downstream of the source were reduced by between 20 and 40 dB. Their experiments show levels of attenuation at least 10 dB higher than predicted, whereas simulations typically provide an estimate of the maximum levels of attenuation one could expect to achieve.

The reason for this discrepancy can be found by considering the traveling wave model. It can be shown that the solution to minimizing the pressure at two microphones downstream of the source results in complete attenuation of pressure at all locations downstream of the control source.[24] It therefore follows that the necessary relationship between the primary source of strength q_p located at $z=0$ and the secondary source at z_s is

$$q_s=-q_p\frac{e^{-jkz_s}(1+e^{2\Phi_1})}{e^{-2jkz_s}+e^{2\Phi_1}}, \tag{6}$$

which converges to the solution obtained for an infinite waveguide as $\Re\{\Phi_1\}\to\infty$, namely $q_s=-q_pe^{-jkz_s}$. The pressure response after control is therefore given by

$$p(z,z_s)=\begin{cases}-\dfrac{\rho c}{2S}\dfrac{e^{-jkz}(e^{2jkz}-e^{2jkz_s})(1+e^{2\Phi_1})}{1+e^{2jkz_s+2\Phi_1}} & z\leqslant z_s\\ 0 & z\geqslant z_s\end{cases}. \tag{7}$$

The net effect of the control source is to completely reflect the sound back upstream via a pressure release boundary condition.[24] The reason that this solution was not achieved in the previous work is due to small phase errors in the pressure response estimates. Modal truncation is the source of these errors in the numerical modal model. In fact, as the total number of modes in the formulation is increased, the solution converges to the traveling wave solution up to a point, as illustrated in Fig. 2. Eventually the numerical precision of the computer program is reached, which then introduces noise and prevents any further convergence. This has been observed by the current authors. The noise in the previous experiments may have come from many sources[25] such as resolution bandwidth errors, poor coherence, etc.

The closed circular duct used in Refs. 1–4 has been used to illustrate these results and issues. Figure 2 shows the pressure response of a duct $L=5.6$ m in length and 0.116 m in diameter, driven at a frequency of 200 Hz by a primary source positioned at one end of the duct and a control source located arbitrarily at $0.34L$. The speed of sound $c=343$ m/s and the density $\rho=1.2$ kg/m^3. The modal quality factor of $Q=50$ was used, corresponding to termination phasors of approximately $\Phi_1=\Phi_2\approx 1/9$. Two microphones separated by 1.7 cm with a midpoint located at $0.47L$ were used as error sensors. The results from the traveling wave solution are compared with results from a modal model using various numbers of modes. The primary sound field for all solutions is almost identical, as is the controlled sound field upstream of the control source. However, the controlled pressure downstream of the control source is quite different, even between modal solutions. Note that the controlled sound field of the modal model begins to converge to the traveling wave solution as more modes are added. Eventually, the modal model ceases to change with the addition of modes. It was found that this is due to numerical noise associated with the double-precision calculations.

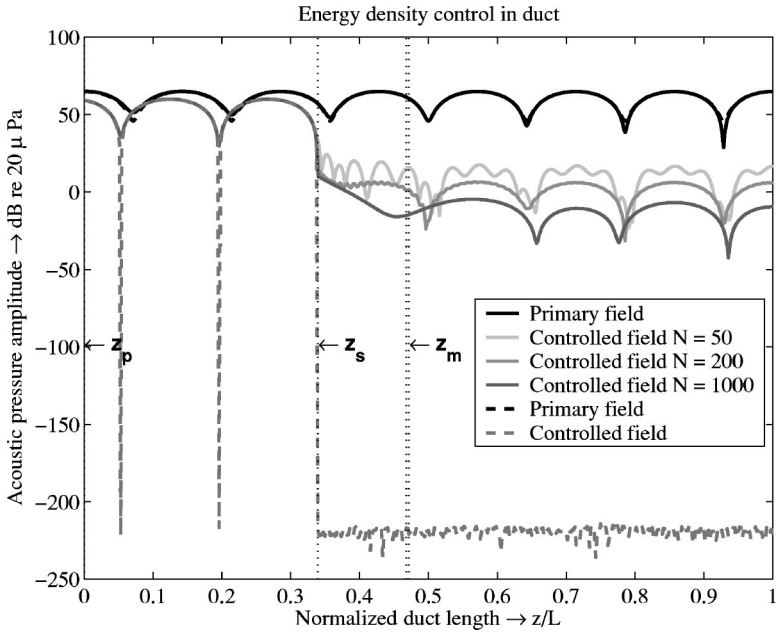


FIG. 2. Pressure response in a duct with a single control source (z_s) located upstream of two microphones (z_m). The primary source was located at one end of the duct (z_p). The black lines indicate the primary sound field and the gray lines indicate the controlled field. The solid lines have been generated using the modal model and the dashed lines are using the traveling wave model. The parameter N represents the number of axial modes included in the modal model.

B. Two control sources

Since it was shown in Sec. III A that only a single upstream control source is required to attenuate completely the acoustic energy density downstream of the control source, it is clear that two control sources upstream of the two microphones will result in a rank-deficient control matrix.

Kestell *et al.*[5–8] investigated the case of two control sources downstream of the two microphones. It will be shown in the following analysis that this is a poorly posed problem and that the control matrix is rank deficient.

The transfer function between two microphones located at z_{m_1} and z_{m_2} when driven by a source located at z_s in the duct is given by

$$H_{12}=\begin{cases}\dfrac{1+e^{2jkz_{m_2}+2\Phi_1}}{e^{-jk(z_{m_1}-z_{m_2})}(1+e^{2jkz_{m_1}+2\Phi_1})} & 0\leqslant z_{m_1}<z_{m_2}<z_s\leqslant L,\\ \dfrac{(1+e^{2jkz_s+2\Phi_1})(e^{2jkz_{m_2}}+e^{2jkL+2\Phi_2})}{e^{-jk(z_{m_1}-z_{m_2})}(1+e^{2jkz_{m_1}+2\Phi_1})(e^{2jkz_s}+e^{2jkL+2\Phi_2})} & 0\leqslant z_{m_1}<z_s<z_{m_2}\leqslant L,\\ \dfrac{e^{2jkz_{m_2}}+e^{2jkL+2\Phi_2}}{e^{-jk(z_{m_1}-z_{m_2})}(e^{2jkz_{m_1}}+e^{2jkL+2\Phi_2})} & 0\leqslant z_s<z_{m_1}<z_{m_2}\leqslant L.\end{cases} \tag{8}$$

From Eq. (8) it can be seen that, when the source is not positioned between the two microphones, the transfer function between the two microphones is independent of the source position. This is to be expected and is the basis for the two-microphone transfer function method used to measure impedance and absorption in impedance tubes.[26] In fact, solving Eq. (8) for the complex reflection coefficient when the source is upstream of the two microphones gives

$$R_2=e^{-2\Phi_2}=e^{2jk(L-z_{m_1})}\left(\frac{H_{12}-e^{-jks}}{e^{jks}-H_{12}}\right), \tag{9}$$

where $s=z_{m_2}-z_{m_1}$ is the microphone separation. This is the same expression derived by Chung and Blaser.[26]

Given that the transfer function between the two microphones is independent of source position (z_s) when the source is located downstream (or upstream) of both microphones, it is then obvious that the second control source does not offer any more control authority than a single control source. Therefore, two control sources located downstream (or upstream) of an energy density sensor is a rank-deficient problem. This implies that both the numerical and experimental work of Kestell *et al.*[5–8] is incorrect and explains why they had such poor correlation between their numerical simulations and experimental results.

Presumably the reason that they were able to calculate the strengths of the control sources at all was due to the same reasons discussed in the previous section. Noise in both the numerical and experimental results meant that the condition number of the control matrix was no longer infinite and therefore invertible. However, this matrix would have been extremely poorly conditioned, indicative of a poorly posed

control problem. This has in fact been observed in simulations using modal models conducted by the current authors.

IV. CONCLUSIONS

It has been shown that, for a single control source in a finite rigid-walled one-dimensional waveguide, it is possible to achieve infinite attenuation of energy density when the primary and control source is located upstream of the energy density sensor. This explains the discrepancy Sommerfeldt *et al.*[1–4] observed between numerical and experimental results. It has also been shown that adding a second control source downstream of the sensor results in a rank-deficient control matrix. This implies that all the previous simulations and experiments investigating the zone of local control using energy density sensing by Kestell *et al.*[5–8] are incorrect, and the conclusions drawn from these results may be flawed. It is perhaps worth reinforcing that the results presented here are not exclusively limited to 1D waveguides, and that results for 3D simulations also need to be checked for modal convergence.

ACKNOWLEDGMENT

The financial support of the Australian Research Council is gratefully acknowledged.

APPENDIX: ENERGY DENSITY CONTROL

The instantaneous acoustic energy density at a point z is given by[1]

$$E_D(z)=\frac{p^2(z)}{2\rho c^2}+\frac{\rho v^2(z)}{2}, \tag{A1}$$

where p and v are the instantaneous acoustic pressure and particle velocity at z, respectively. The particle velocity is related to the pressure gradient by the following:

$$v(z)=-\frac{1}{\rho}\int\frac{\partial p(z)}{\partial z}dt. \tag{A2}$$

The spatial derivative of the pressure is commonly estimated using two closely spaced microphones to provide a first-order spatial derivative given by

$$\frac{\partial p(z)}{\partial z}\approx\frac{p_2-p_1}{2h}, \tag{A3}$$

where p_1 and p_2 are the pressures at the two microphones separated by a distance of $2h$. Evaluating the temporal integral in Eq. (A2) and substituting the pressure gradient estimate given by Eq. (A3), the particle velocity estimate can be rewritten in terms of the acoustic wave number $k=\omega/c$, and is given by

$$v(z)=\frac{j}{\rho\omega}\frac{p_2-p_1}{2h}=\frac{j}{\rho ck}\frac{p_2-p_1}{2h}. \tag{A4}$$

Using the average of the two microphone signals as the pressure estimate, the instantaneous acoustic energy density at the midpoint between points 1 and 2 is given by

$$\hat{E}_D(z)\approx\frac{1}{8\rho c^2}\left[|(p_1+p_2)|^2+\frac{|(p_2-p_1)|^2}{(kh)^2}\right]. \tag{A5}$$

[1] P. J. Nashif and S. D. Sommerfeldt, "An active control strategy for minimizing the energy density in enclosures," in Proceedings of Inter Noise 92, 1992, pp. 357–361.

[2] S. D. Sommerfeldt and P. J. Nashif, "Energy based control of the sound field in enclosures," in The Second International Congress on Recent Developments is Air- and Structure-Borne Sound and Vibration, 1992, pp. 361–368.

[3] S. D. Sommerfeldt and P. J. Nashif, "An adaptive filtered-x algorithm for energy based active control," J. Acoust. Soc. Am. **96(1)**, 300–306 (1994).

[4] Y. C. Park and S. D. Sommerfeldt, "Global attenuation of broadband noise fields using energy density control," J. Acoust. Soc. Am. **101(1)**, 350–359 (1997).

[5] C. D. Kestell, B. S. Cazzolato, and C. H. Hansen, "Virtual energy density sensing in active noise control systems," in Proceedings of the International Congress on Sound and Vibration, 2000.

[6] C. D. Kestell, C. H. Hansen, and B. S. Cazzolato, "Active noise control with virtual sensors in a long narrow duct," Int. J. Acous. Vibr. **5(2)**, 63–76 (2000).

[7] C. D. Kestell, C. H. Hansen, and B. S. Cazzolato, "Virtual sensors in active noise control," Acoustics Australia **29(2)**, 57–61 (2001).

[8] C. D. Kestell and C. H. Hansen, "Active noise control with virtual sensors," in Proceedings of the International Congress on Sound and Vibration, 2001.

[9] B. S. Cazzolato and C. H. Hansen, "Errors in the measurement of acoustic energy density in one-dimensional sound fields," J. Sound Vib. **236(5)**, 801–831 (2000).

[10] S. D. Sommerfeldt, J. W. Parkins, and Y. C. Park, "Global active noise control in rectangular enclosures," in Proceedings of Active 95, 1995, pp. 477–488.

[11] S. J. Elliott and J. Garcia-Bonito, "Active cancellation of pressure and pressure gradient in a diffuse sound field," J. Sound Vib. **186(4)**, 696–704 (1995).

[12] S. D. Sommerfeldt, J. W. Parkins, and J. Tichy, "Modal results for active control of energy density in a rectangular enclosure," in Proceedings of Noise-Con 96, 1996, pp. 429–434.

[13] W. Shen and J. Q. Sun, "A study of shell interior noise control," Proc. SPIE **3041**, 812–818 (1997).

[14] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, "Error analysis of a practical energy density sensor," J. Acoust. Soc. Am. **108(1)**, 211–222 (2000).

[15] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, "Narrow-band and broadband active control in an enclosure using the acoustic energy density," J. Acoust. Soc. Am. **108(1)**, 192–203 (2000).

[16] S. K. Lau and S. K. Tang, "Sound fields in a slightly damped rectangular enclosure under active control," J. Sound Vib. **238(4)**, 637–660 (2000).

[17] B. S. Cazzolato and C. H. Hansen, "Errors arising from three-dimensional acoustic energy density sensing in one-dimensional sound fields," J. Sound Vib. **236(3)**, 375–400 (2000).

[18] S. K. Lau and S. K. Tang, "Impacts of structural-acoustic coupling on the performance of energy density-based active sound transmission control," J. Sound Vib. **266(1)**, 147–170 (2003).

[19] B. S. Cazzolato and C. H. Hansen, "Active control of enclosed sound fields using three-axis energy density sensors: Rigid walled enclosures," Int. J. Acous. Vib. **8(1)**, 39–51 (2003).

[20] D. Lubman, "Equivalence of eigenmode and free-wave models of steady-state reverberation in rectangular rooms," J. Acoust. Soc. Am. **60(Supp 1)**, S59 (1976).

[21] A. C. Zander and C. H. Hansen, "A comparison of error sensor strategies for the active control of duct noise," J. Acoust. Soc. Am. **94(2)**, 841–848 (1993).

[22] S. J. Elliott and P. A. Nelson, "Models for describing active noise control in ducts," ISVR Technical Report 127, ISVR, University of Southampton, 1984.

[23] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[24] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).

[25] J. S. Bendat and A. G. Piersol, *Random Data—Analysis and Measurement Procedures*, 2nd ed (Wiley, New York, 1986).

[26] J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68(3)**, 907–913 (1980).

Tracking the time to recovery after induced loudness reduction (L)

Yoav Arieh[a)]
The Psychology Department, Montclair State University, Montclair, New Jersey 07043

Karen Kelly
The Psychology Department, Montclair State University, Montclair, New Jersey 07043

Lawrence E. Marks
John B. Pierce Laboratory and Yale University, New Haven, Connecticut 06519

(Received 31 December 2004; revised 4 March 2005; accepted 4 March 2005)

In induced loudness reduction (ILR), a strong tone causes the loudness of a subsequently presented weak tone to decrease. The aim of the experiment was to determine the time required for loudness to return to its initial level after ILR. Twenty-four subjects were exposed to 5, 10, 20, or 40 brief bursts of 2500-Hz pure tones at 80-dB SPL (inducers) and then tested in a series of paired comparison trials. Subjects compared the loudness of a weak target (2500 Hz at 60-dB SPL) to the loudness of a comparison tone at 500 Hz previously judged to match the target. The comparison task was repeated until the two tones were again judged equally loud. The results showed that (a) recovery after ILR is a relatively long process with a time scale of minutes, and (b) recovery time increased approximately 20 s with each doubling of the number of inducers. [DOI: 10.1121/1.1898103]

PACS numbers: 43.66.Cb [AJO] Pages: 3381–3384

I. INTRODUCTION

Under appropriate conditions, presenting a relatively strong inducing tone reduces the loudness of a subsequent weaker tone (Arieh and Marks, 2003a; Mapes-Riordan and Yost, 1999; Marks, 1994; Nieder *et al.*, 2003). The phenomenon was initially called recalibration (Marks, 1994), on the premise that it reflects a general principle of intensity processing in the nervous system. Later, the name induced loudness reduction (ILR) was offered to describe the effect specifically in hearing (Scharf *et al.*, 2002). Because this report deals exclusively with loudness, we shall adopt the latter term here.

Four important properties of ILR have been determined since it was first reported (Marks, 1988). First, under optimal conditions the extent of the loudness reduction can reach 10 dB or more (Arieh and Marks, 2003a; Nieder *et al.*, 2003). In other words, a stronger inducer can reduce the loudness, in sones, of the weak target by at least half. Second, ILR is frequency specific, being greatest when the inducer and the target fall within the same critical band (Marks and Warner, 1991). Third, the inducer must precede the target by at least 200 ms (offset–onset) for significant ILR to appear (Arieh and Marks, 2003b). And fourth, ILR is maximal when the SPL of the inducer is 60–80 dB SPL and the inducer and target differ in level by 10–20 dB (Mapes-Riordan and Yost, 1999).

Importantly, ILR does not reflect a response bias but instead is a sensory change, most likely a change in the sensory representation of the intensity of the target tone. Arieh and Marks (2003b) showed that when listeners detect weak tones in a choice decision task, conditions that produce ILR also produce longer response times and higher error rates than do control conditions. The positive relation between response time and error rate is a hallmark of sensory as opposed to decisional change (Luce, 1986).

Still lacking, however, are data about recovery after ILR. After ILR, how long does it take for the loudness of the target tone to return to its original level? The purpose of the current report was to systematically explore this question.

The literature does offer a few clues about recovery after ILR. Arieh and Marks (2003a) showed that 3.3 s after the presentation of the inducer, the loudness of the target tone is still reduced substantially, such that the level of the matching tone is 13 dB less than the matched level without the inducer. Mapes-Riordan and Yost (1998) reported that, even when the interval between the inducer and the target tone is increased to 10 s, ILR remains strong. Marks (1993) noted that ILR largely dissipates when a 60-s pause follows a condition in which ILR is induced. Thus, while ILR arises quickly, within 200 ms, recovery after ILR is a relatively long process requiring some dozens of seconds.

The present study also examines how recovery from ILR depends on the number of inducers. It is reasonable to expect, for instance, that recovery time after ILR will depend on the magnitude of ILR itself: the greater the reduction in loudness, the greater the time to recover. Systematic data about the way magnitude of ILR depends on number of inducers are also lacking. In many studies of ILR, the number of inducers was not controlled nor even reported. For example, in an adaptive procedure often used to study ILR (Arieh and Marks, 2003a; Mapes-Riordan and Yost, 1999), the matching points depend on the listeners' patterns of responses, so the number of trials (equal to the number of

a)Electronic mail: ariehy@mail.montclair.edu

presentations of the inducer) can vary considerably between listeners and between experimental conditions.

The effect, if any, of number of inducers bears importantly on the nature of ILR. It is possible, for example, that ILR is an ''all or none'' process that begins anew with each presentation of an inducer. If so, then varying the number of inducers would have no effect on the magnitude of ILR, or on the time for recovery. Alternatively, like so many processes of ''sensory adaptation,'' ILR may reflect the outcome of a suppressive process that builds up through repeated presentation of inducers, and hence requires increasingly greater time to dissipate.

A word is in order as to the logic of the study. Subjects started an experimental session by listening to uninterrupted series of 5, 10, 20, or 40 inducers. Immediately thereafter subjects were presented with a sequence of paired comparison trials. Each trial consisted of the target tone with the same frequency as the inducers and a comparison tone with a different frequency. The latter was previously deemed equal in loudness to the target tone. The subject's task was to judge which tone was louder. We reasoned that if the target tone underwent ILR by the preceding inducers, the listeners would tend to choose the comparison tone as louder. However, as time elapsed ILR would dissipate and loudness would recover. At this stage listeners would increasingly choose the target tone as the louder of the pair. According to this procedure, when the target tone regains its original loudness, listeners' performance would be at or close to chance level. Thus, the point of full recovery was defined as the point where subjects performed the paired comparison task at chance level.

II. RECOVERY FROM ILR AS A FUNCTION OF THE NUMBER OF INDUCERS

A. Method

Twenty women and four men (age range 18–40), all associated with Montclair State University, participated in the study. All reported normal hearing.

Testing sessions were conducted in a sound-treated chamber. A MATLAB program running on a Dell Pentium-IV PC controlled all aspects of stimulus presentation, data collection, and on-line computations. Subjects' responses were entered via the computer keypad. The pure-tone signals were generated by a Tucker-Davis System 3 real time processor at a sampling frequency of 50 kHz. The signals were then appropriately attenuated (Tucker-Davis PA5 module) and delivered binaurally for 50 ms (including 5-ms cosine^2 rise and decay) through calibrated TDH-49 headphones mounted in MX41/AR cushions.

Before the start of each experimental session, the subject performed a baseline task to determine the matching point between the comparison and the target tone. It is critical to establish a reliable matching point because this point serves later to gauge recovery time. In the baseline task, we used a randomized adaptive two-track ascending and descending procedure to determine the level of a 500-Hz tone equal in loudness to the designated target signal—a 2500-Hz tone at 60-dB SPL (details of the procedure can be found in Jesteadt, 1980, and Arieh and Marks, 2003a).

After determining the matching point, the subject served in four experimental conditions divided into two different sessions separated by at least 24 h. Each condition was divided in turn into two continuous parts. In the first part, the subject heard an uninterrupted series of 5, 10, 20, or 40 2500-Hz tones at 80-dB SPL that served as inducers. The 50-ms inducing tones were presented at intervals of 1 s (offset–onset). On the screen, a visual countdown accompanied the inducer sequence and upon its termination a pair of tones was presented for paired comparison. The two tones were the 500-Hz matching level determined at baseline and the 2500-Hz, 60-dB target. The subjects were instructed to judge which tone was louder as quickly as possible without sacrificing accuracy. The order of the two tones was randomized across trials and the tones were presented 1 s apart. The program recorded the subject's response and its latency from the onset of the paired comparison trial.

After the first ten trials the probability p of selecting the target tone as louder was computed. If p fell between 0.4 and 0.6, it was concluded that the subject performed at chance and full recovery had been reached (a final p larger than 0.6 was also considered as complete recovery; however, none occurred in the current study). On the other hand, if p was smaller than 0.4, it was concluded that recovery was not complete and another trial was presented. From this point, the probability of selecting the target tone as louder was continually computed over the ten most recent trials until full recovery had been reached, after which the session terminated.

It was assumed that recovery occurred within the time period that elapsed between the first and the last of the final ten trials. Thus, the midpoint of that time period was taken as the recovery time. For example, if recovery occurred between trial numbers 32 and 41 and these trials occurred 90 and 120 s, respectively, after the start of session, then recovery time was recorded as 105 s. Finally, to minimize the possibility of order effects, we used all 24 possible orders of the four inducer conditions. Each subject received one possible order; the first two conditions were performed in the first session and the last two conditions were performed in the second session.

B. Results

As anticipated, there were no effects of order. The correlation between the recovery time and the serial position of a condition in the presentation order was negligible, $r(94) = -0.039$, $p = 0.78$. The average proportion of selecting the target tone over the first ten trials was 0.229, 0.194, 0.180, and 0.138 and over the last ten trials was 0.421, 0.408, 0.420, and 0.413 for the conditions containing 5, 10, 20, and 40 inducers, respectively. A repeated measures analysis of variance revealed that the final averages did not differ statistically ($F < 1$). Given that $p = 0.5$ represents chance performance, recovery was caught a bit early but equally so in all conditions.

The resolution of our measuring method is the average time for making a paired comparison judgment, which was 3

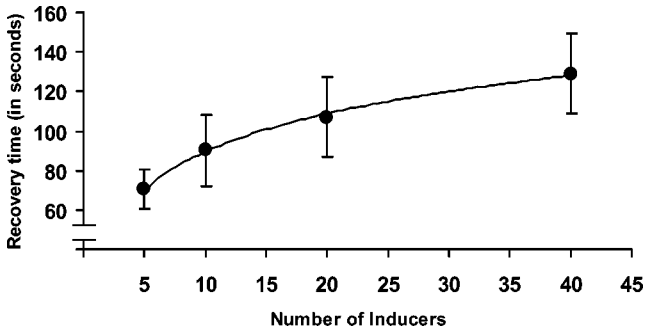


FIG. 1. Averaged recovery times (in seconds) for the 24 listeners plotted against the number of inducers. The error bars represent one standard error of the mean.

s. Thus, the average time to complete the first ten trials that are necessary for the computation of p is 30 s, and the minimal recovery time afforded by our procedure is 15 s. As the data show, these limitations were inconsequential under the stimulus conditions.

The recovery time for each inducer condition was averaged across subjects. The averages are presented in Fig. 1. As can be seen, recovery time increases with the number of inducers. The shortest average recovery time was 70.5 s for the 5-inducer condition and the longest was 129 s for the 40-inducer condition. A repeated-measure analysis of variance showed a main effect of inducer condition, $F(3,69) = 3.64$, $p < 0.05$. A closer look at Fig. 1 also reveals that, within the tested range, the function that relates the number of inducers to recovery time is not linear. In fact, every time the numbers of inducers is doubled an approximately constant amount of time—around 20 s—is added to the recovery time. This suggests that recovery time increases approximately with the logarithm of the number of inducers. The curved line in Fig. 1 shows a log function fitted to the data, computed by regressing recovery time on log number of inducers; the fit is excellent, explaining 99% of the total variance. The small number of data points, however, prohibits a conclusive test of the form of the function. For example, a linear fit explains 93% of the variance in our data set—less than that produced by the logarithmic fit, but still substantial. Thus, at this point any conclusion about the function relating recovery time to number of inducers must be tentative.

One possible limitation of the current procedure is that recovery time is correlated with the number of target presentations. That is, the longer the recovery time the greater the number of comparisons, and, consequently, the greater the number of exposures to the target tones. It is possible that these additional exposures prolong recovery times. To evaluate this possibility we ran ten additional subjects in a control experiment that tested recovery from 40 inducers only, but under two conditions. The first condition replicated the original experiment, in that the paired comparisons began as soon as the induction phase ended. In the second condition, we inserted a 30-s pause between the induction phase and the start of the judgment phase. Thus, in the second condition the subjects were allowed to recover for 30 s before being exposed to the target tones. Consequently, they were exposed, on average, to ten fewer target presentations than the no-pause group. Average recovery times for the no-pause and pause conditions were 113.1 and 119.5 s, respectively, the difference (dependent-pair t-test) not being reliable [$t(9) = 1.2$, ns]. Thus, we conclude that exposure to ten additional target presentations in the no-pause group during recovery had no significant effect on the final recovery time.

III. DISCUSSION

The measurement of recovery after ILR, and indeed the measurement of sensory recovery in general, presents an interesting challenge. Recovery, by definition, is a dynamic process that continuously changes with the passage of time. On the other hand, reliable measurement necessitates several observations of the same event. Observations, especially behavioral observations, however, take time. How can a constantly changing event be measured reliably?

Our solution is to define recovery as a probabilistic process that is computed over a roving window of observations comprising ten successive loudness comparisons. Once the listener can no longer reliably discriminate between the target tone that underwent ILR and a comparison tone that was previously deemed equally loud, recovery has been reached. The method is not perfect, its resolution being about 3 s, and it cannot measure recovery times smaller than 15 s. A crucial consideration is the size of the observation window. Increasing the window's size will increase reliability but will also increase the lower limit of measurement; decreasing the size will decrease the lower limit of measurement but will also decrease reliability. We chose our parameters after careful consideration of the available literature on recovery after ILR. Applying the method to other sensory processes would undoubtedly require adjustments to its parameters.

Our report is the first to measure recovery after ILR to the point where the target's loudness returned to or near to its original level. A main finding is that recovery time increases monotonically with the number of inducer tones presented in the exposure phase. The fastest recovery time, recorded after exposure to five inducers, exceeded 1 min. The slowest recovery time, after exposure to 40 inducers, exceeded 2 min. These results fit previous observations that hinted that recovery after ILR is a long process, on the order of dozens of seconds (Arieh and Marks, 2003a; Mapes-Riordan and Yost, 1998; Marks, 1993). We note here the asymmetry between the onset and the offset of ILR. The onset is a fast process; ILR is already present 200 ms after the presentation of the inducer. The offset has a completely different time scale; dissipation of ILR takes over a minute after exposure to five inducers. Any future model of ILR will have to account for this property.

We also found that a log function provides a good description of the way recovery time depends on the number of inducers. Put simply, adding inducers does not lead to proportional addition of recovery time. In the range used here, each doubling of the number of inducers added another 20 s to recovery time. Regardless of the exact shape of the function that relates the number of inducers to recovery time, the fact that they are positively related is important. Assuming that the magnitude of ILR itself is associated with recovery time, we can surmise that the former is also positively related to the number of inducers. That conclusion supports a model

of ILR that has ''memory'' for previous exposures to inducers: It allows for accumulation of loudness suppression effects over time. On the other hand, the conclusion does not support an ''all or none'' model in which the suppression effect resets on every exposure to inducer. Admittedly, a direct measurement of the relation between the magnitude of ILR and recovery time will go a long way towards clarifying this issue.

That the magnitude of ILR is related to the number of inducers also has practical implications. Many studies used an adaptive procedure to measure ILR (Arieh and Marks, 2003a, Mapes-Riordan and Yost, 1999). In this procedure the inducer is presented on each trial but the number of trials depends on the pattern of responses and thus often varies between listeners and between experimental conditions. The fact that the number of inducers has not been controlled in many earlier studies might contribute to the significant variability that is often observed in studies of ILR.

ACKNOWLEDGMENTS

Preparation of this article was supported in part by Grant DC00271-17 from the National Institutes of Deafness and Other Communication Disorders to L.E.M. We would like to thank Scott Parker, Christopher Plack, and Andrew Oxenham for their cogent comments on the earlier version of this paper.

Arieh, Y., and Marks, L. E. (**2003a**). ''Time course of loudness recalibration: Implications for loudness enhancement,'' J. Acoust. Soc. Am. **114**, 1550–1556.

Arieh, Y., and Marks, L. E. (**2003b**). ''Recalibrating the auditory system: A speed-accuracy analysis of intensity perception,'' J. Exp. Psychol. Hum. Percept. Perform. **29**, 523–536.

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgments,'' Percept. Psychophys. **28**, 85–88.

Luce, R. D. (**1986**). *Response Times* (Oxford University Press, New York).

Mapes-Riordan, D., and Yost, W. A. (**1998**). ''Temporal properties of loudness recalibration,'' Proceedings, 16th Int. Cong. Acoust., pp. 2361–2362.

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. (**1988**). ''Magnitude estimation and sensory matching,'' Percept. Psychophys. **43**, 511–525.

Marks, L. E. (**1993**). ''Contextual processing of multidimensional and unidimensional auditory stimuli,'' J. Exp. Psychol. Hum. Percept. Perform. **19**, 227–249.

Marks, L. E. (**1994**). ''Recalibrating the auditory system: The perception of loudness,'' J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.

Marks, L. E., and Warner, E. (**1991**). ''Slippery context effect and critical bands,'' J. Exp. Psychol. Hum. Percept. Perform. **17**, 986–996.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). ''Interactions between test-and inducer-tone durations in induced loudness reduction,'' J. Acoust. Soc. Am. **114**, 2846–2855.

Scharf, B., Buus, S., and Nieder, B. (**2002**). ''Loudness enhancement: induced loudness reduction in disguise? (L),'' J. Acoust. Soc. Am. **112**, 807–810.

Effects of vocalic duration and first formant offset on final voicing judgments by children and adults (L)

Caroline Jones
School of Education, The University of New South Wales, Sydney New South Wales 2052, Australia and MARCS Auditory Laboratories, University of Western Sydney, Penrith South 1797, New South Wales, Australia

(Received 12 January 2005; revised 15 March 2005; accepted 22 March 2005)

Developmental research into relative weighting of vocalic duration and spectral properties in final voicing perception has produced different results, depending on whether natural or synthetic speech stimuli have been used and how spectral properties have been manipulated. This paper reports developmental data for final voicing using natural stimuli waveform edited for vocalic duration and resynthesized for first formant offset. Results indicate that in perception of final voicing there are adult-child differences in weighting of vocalic duration and first formant offset, consistent with previous findings for vocalic duration and spectral properties more generally [S. Nittrouer, J. Acoust. Soc. Am. **115**, 1777–1790 (2004)]. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1906058]

PACS numbers: 43.71.Ft, 43.71.An [RLD] Pages: 3385–3388

I. INTRODUCTION

There is increasing evidence that preschool age children perceive speech on the basis of multiple acoustic properties, but weight these properties differently from adults (Allen and Norwood, 1988; Gerrits, 2001; Mayo and Turk, 2004; Morrongiello *et al.*, 1984; Nittrouer, 1992; Nittrouer *et al.*, 1988; Nittrouer and Studdert-Kennedy, 1987). For example, in a study of the English prevocalic fricative contrast *sigh/shy* (Mayo and Turk, 2004), children aged 3–4, 5, and 7 years weighted formant transitions more heavily than adults.

Final voicing in English has also received attention in research into children's speech perception, but little is known about the relative weighting of different properties by children. Greenlee (1980) and Krause (1982) manipulated preceding vowel duration in natural and synthetic stimuli, respectively, and the results suggest that 3- and 6-year-old children can use vowel duration alone in identifying final voicing if the voicing-related vowel duration differences are large enough. Wardrip-Fruin and Peach (1984) aimed to assess 3- and 6-year-old children's use of vowel duration, final transition, and closure voicing in identifying final voicing, but the variability in the results and the small scale of the study make it difficult to draw conclusions about the specific effects of these properties on children's judgments. Lehman and Sharf (1989) tested voicing identification and discrimination by 5-, 8-, and 10-year-old children with synthetic stimuli varying in vocalic duration, and found decreasing variability in identification and improved discrimination with increasing age.

With respect to the relative weighting of acoustic properties by children, one proposal is the Developmental Weighting shift (Nittrouer *et al.* 1993; Nittrouer and Miller, 1997): the hypothesis that children weight formant transitions more heavily than adults, which is based on the idea that children process speech more globally than adults, in larger units such as syllables. This hypothesis has received some support in experimental studies of children's speech perception, mainly studies of prevocalic fricative perception (e.g., Nittrouer, 1992; Nittrouer and Miller, 1997), but some research has found clear evidence against it (e.g., Gerrits, 2001; Mayo and Turk, 2004).

In recent research designed to test the Developmental Weighting shift hypothesis, Nittrouer (2004) tested children's perception of final voicing in American English using synthetic and natural speech stimuli. For synthetic *buck/bug* and *pot/pod* stimuli, which varied in vocalic duration (nine steps) and first formant (F1) offset (two levels), 6- and 8-year-olds weighted F1 offset more and vocalic duration less, compared with adults, as measured by partial correlation coefficients (partial *r*'s) and function slope and separation. Natural *buck/bug* stimuli were constructed on the basis of three natural tokens of each of *buck* and *bug*, with release burst and closure voicing deleted. Vocalic duration was manipulated by deleting pitch periods, or reiterating a single pitch period, from the vowel steady state. The two series based on original *buck* and *bug* recordings differed in fundamental frequency (F0) and all formants (although recordings which matched closely in F0 were selected as the basis for the stimuli).

As compared with synthetic stimuli, for natural stimuli both adult and child judgments were heavily influenced by the spectral variation, and manipulations of vocalic duration had correspondingly less perceptual effect. Differences between adults and children in weighting of vocalic duration were smaller (by slope measure) or nonexistent (by partial *r* measure). Spectral variation influenced 6-year-olds more than adults (by partial *r*, but not by the separation in functions).

A follow-up experiment using natural stimuli with vocalic durations comparable to the synthetic stimuli [Experiment 2 in Nittrouer (2004)] found no age-related differences in relative weighting of vocalic duration and spectral properties, except that adults were more likely to give voiced judgments to stimuli with longer vocalic durations if they were also derived from originally voiced tokens. On the ba-

sis of results from natural stimuli, Nittrouer (2004) concluded that the strong effect of spectral properties on voicing judgments was most likely due to syllable-offset transitions, especially F1. Nittrouer (2004) also noted, however, that the natural stimuli also varied in higher formant transitions, and formant transitions corresponding to voiced and voiceless variation were present in syllable onset and steady state (when the latter was not shortened in stimulus construction).

Given the difference in results between synthetic and natural stimuli in Nittrouer (2004) as well as the range of results in previous research on children's perception of final voicing (Greenlee, 1980; Krause, 1982; Lehman and Sharf, 1989; Wardrip-Fruin and Peach, 1984), the relative influence of vocalic duration and F1 offset transitions on final voicing judgments by children and adults warrants further investigation. In particular, in the natural modified stimuli in Nittrouer (2004), F1 variation between voiced and voiceless offset transitions was possibly confounded with variation in F0, F2, and F3. From the data in Nittrouer (2004) alone, it is not possible to conclude for certain that the results are due to variation in F1, and not variation in F0, F2, or F3. This paper provides specific data on the relative weighting of vocalic duration and F1 offset in judgments of final voicing in English by children and adults, using modified natural speech stimuli in which F1 alone was manipulated and all other spectral properties were held constant. The dialect studied is Australian English; these data complement existing studies of final voicing in American English.

II. METHOD

A. Subjects

The participants were 19 adults (mean age 20 years, range 18 to 40) and 19 children (mean age 4 years 10 months, range 4 years 2 months to 5 years 6 months). Children were recruited from preschools in Sydney; adults were 18 first-year undergraduate students at the University of Western Sydney participating for course credit in an introductory psychology course and one graduate student in psychology who was a volunteer. No participants had phonetic training. Male/female ratios were 1/18 for adults and 8/11 for children. According to self-report (adults) and parental report (children), all participants had normal hearing and speech, and no history of past or present recurrent middle ear infections. Data from one additional adult participant and three additional child participants are not included here. These listeners did not meet 70% accuracy during testing on the two endpoint stimuli (*hard*: longest vocalic duration, low F1 offset; *heart*: shortest vocalic duration, high F1 offset).

B. Stimuli

A recording of *hard* spoken by a young male speaker of Australian English was modified to produce a set of 14 *heart-hard* stimuli varying in vocalic duration (seven steps, 126 to 293 ms, step size was 26 or 27 ms except for the last step which was 33 ms) and F1 offset frequency (two levels: falling linearly from 770 to 650 or 530 Hz over the last 50 ms of the vowel). The release burst and closure voicing were deleted. Vocalic duration was varied by successively deleting pitch periods from the steady-state portion of the vowel. F1 offset manipulations were made by resynthesizing the last 50 ms of the vowel (using Kay CSL LPC Parameter Manipulation/Synthesis Program, Model 4304, software version 1.X). All 14 stimuli had the same F0, F2, and F3 values (from the original *hard* recording) listed in Table I. The stimuli were audibly and visually free of artefacts from the resynthesis and waveform editing. The wordform log frequency in the CELEX English database (version 2.5) is similar for *heart* (2.1614) and *hard* (2.2672).

TABLE I. Fundamental frequency and formant values for the stimuli (in Hz).

	Time vowel onset	50 ms before vowel offset	Vowel offset
F0	139	108	101
F1 (low F1 series)	784	770	530
F1 (high F1 series)	784	770	650
F2	1427	1349	1358
F3	2690	2758	2906

C. Procedures

Participants were tested individually. Each participant judged 11 repetitions of each of the 14 stimuli, for a total of 154 judgments per listener. The experiment was run on a PC laptop (IBM Thinkpad R31) with high-quality external speakers (Cambridge SoundWords PC Works), and programmed in game format using DMDX. (DMDX is programmed by Jonathan Forster, University of Arizona.) Response keys were picture-labeled. Adults were tested in one four-block session in a quiet lab room. Children were tested in four one-block sessions in a quiet room in their preschool. In each block, the participant received four demonstration trials of the two endpoint stimuli (*hard*: 293-ms vocalic duration, low F1 offset; *heart*: 126-ms vocalic duration, high F1 offset), with feedback. Participants completed training trials with the endpoint stimuli, with feedback, until they gave six consecutive correct responses. Test trials did not include feedback. Test block 1 comprised 28 trials; the other test blocks had 42 trials each. Children received stickers and encouragement after each 7 or 11 test trials (depending on the block), and received two endpoint reminder trials before resuming testing.

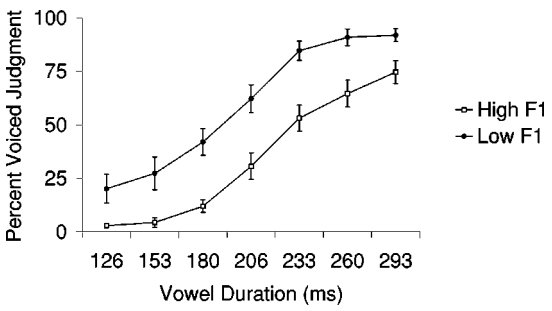


FIG. 1. Mean labeling functions (with standard errors) for adults.

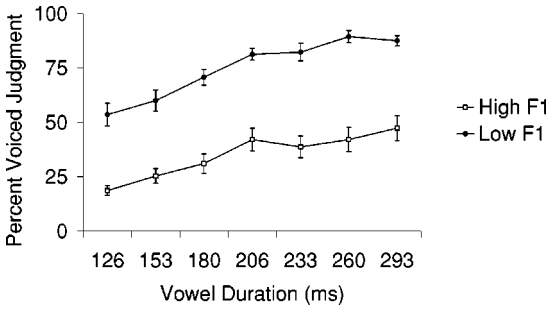


FIG. 2. Mean labeling functions (with standard errors) for 5-year-olds.

III. RESULTS

Figure 1 shows mean labeling functions for adults, Fig. 2 for 5-year-olds. Table II provides estimates of the weighting of vocalic duration [mean slopes, partial correlation coefficients (partial *r*'s) for vocalic duration] and the weighting of F1 offset (separation in phoneme boundaries in steps of vocalic duration, partial *r*'s for F1 offset). Probit analysis was used to calculate 50% crossover points (to measure separation in phoneme boundaries) and slopes. To facilitate comparison with data in Nittrouer (2004), slope is stated as the change in probit units per step on the duration continuum, and the same restrictions as in Nittrouer (2004) are used for 50% crossover estimates (± 3.5 steps beyond the continuum).

Estimates using slope and separation between functions indicate that adults and children differ in the relative weighting of both vocalic duration and F1 offset. The slope estimates indicate that adults weight vocalic duration more than children, $t(22.25) = 4.879$, $p < 0.001$. The separations between the functions indicate that children weight F1 offset more than adults, $t(26) = -3.38$, $p = 0.002$.

Estimates using partial *r*'s present similar patterns. Adults weight vocalic duration more than children [$t(36) = 6.26$, $p < 0.001$]. There is no significant difference between adults and children in the weighting of F1 offset [$t(36) = -1.51$, $p = 0.14$] although there is a tendency for children to weight F1 offset more than adults.

IV. DISCUSSION

These data provide evidence from modified natural speech stimuli that in the perception of final voicing, adults and children differ in the relative weighting of vocalic duration and F1 offset. Adults weight vocalic duration more than children. Children weight F1 offset more than adults when separation in phoneme boundaries is used as an estimate of F1 offset weighting. There is a nonsignificant tendency towards the same effect when partial *r* is used to estimate F1 offset weighting.

The pattern of results obtained here for Australian English is similar to that found in Nittrouer (2004) for American English. The stimuli used in the present study, however, varied in F1 and vocalic duration only, rather than varying in vocalic duration, F1, and possibly other spectral properties as well, as in the stimuli in Nittrouer (2004). The present study thus provides perceptual data for natural stimuli about the specific role of F1 offset in final voicing judgments. The present data suggest that the patterns found for natural stimuli by Nittrouer (2004), where variation in vocalic duration did not compensate perceptually for variation in spectral properties, probably reflect the effects of any or all of the following properties: higher formants, F0, and F1 steady state and onset, as well as F1 offset.

In conjunction with the data from synthetic stimuli in Crowther and Mann (1992) and Nittrouer (2004), the present data suggest that for adults, vocalic duration is an important cue for the perception of final voicing in English which is not generally overridden by variation in F1 offset alone. Children's perception of final voicing, however, appears to develop in childhood, as they come to give more weight to vocalic duration and correspondingly less to F1 offset.

TABLE II. Mean slopes, separations between functions at phoneme boundaries, and partial *r*'s for vocalic duration and F1 offset. Standard deviations are given in parentheses.

	Adults	5-year-olds
Weight given to vocalic duration		
Mean slope, across functions	0.65	0.19
	(0.39)	(0.13)
Mean partial *r* for vocalic duration	0.59	0.24
	(0.20)	(0.15)
Weight given to F1 offset		
Mean separation between functions	1.53	4.96
	(2.46)	(2.87)
Mean partial *r* for F1 offset	0.33	0.42
	(0.22)	(0.16)

ACKNOWLEDGMENTS

The author thanks the child participants, their parents, and preschools for their time and effort.

Allen, G. D., and Norwood, J. A. (**1988**). "Cues for intervocalic /t/ and /d/ in children and adults," J. Acoust. Soc. Am. **84**, 868–875.

Crowther, C. S., and Mann, V. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Gerrits, E. (**2001**). "The categorisation of speech sounds by adults and children: A study of the categorical perception hypothesis and the developmental weighting of acoustic speech cues," Ph.D. thesis, University of Utrecht, http://www.library.uu.nl/digiarchief/dip/diss/1946869/inhoud.htm.

Greenlee, M. (**1980**). "Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception," J. Child Lang **7**, 459–468.

Krause, S. E. (**1982**). "Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults," J. Acoust. Soc. Am. **71**, 990–995.

Lehman, M. E., and Sharf, D. J. (**1989**). "Perception/production relationships in the development of the vowel duration cue to final consonant voicing," J. Speech Hear. Res. **32**, 803–815.

Mayo, C., and Turk, A. (**2004**). "Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," J. Acoust. Soc. Am. **115**, 3184–3194.

Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). "Trading relations in the perception of speech by 5-year-old children," J. Exp. Child Psychol. **37**, 231–250.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 1–32.

Nittrouer, S. (**2004**). ''The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults,'' J. Acoust. Soc. Am. **115**, 1777–1790.
Nittrouer, S., and Miller, M. E. (**1997**). ''Predicting developmental shifts in perceptual weighting schemes,'' J. Acoust. Soc. Am. **101**, 2253–2266.
Nittrouer, S., and Studdert-Kennedy, M. (**1987**). ''The role of coarticulatory effects in the perception of fricatives by children and adults,'' J. Speech Hear. Res. **30**, 319–329.
Nittrouer, S., Crowther, C. S., and Miller, M. E. (**1998**). ''The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults,'' Percept. Psychophys. **60**, 51–64.
Nittrouer, S., Manning, C., and Meyer, G. (**1993**). ''The perceptual weighting of acoustic cues changes with linguistic experience,'' J. Acoust. Soc. Am. **94**, S1865.
Wardrip-Fruin, C., and Peach, S. (**1984**). ''Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants,'' Lang Speech **27**, 367–379.

A note on reflection of spherical waves

Gunnar Taraldsen[a)]
Acoustic Research Center, Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), O.S. Bragstads Plass, N-7465 Trondheim, Norway

(Received 9 November 2004; revised 9 March 2004; accepted 15 March 2004)

In 1909 Sommerfeld gave an exact solution for the reflection of a spherical wave from a plane surface in terms of an oscillatory integral and also presented an asymptotic solution for the case where both source and receiver are at the boundary. Weyl (1919) presented an alternative solution and also an asymptotic solution for the case where the source is at the boundary. It is known that the general case is solved if a general solution for the case where the source is at the boundary is known. Here it is demonstrated that it is sufficient to have the general solution for the case where both source and receiver are at the boundary. This is mainly of theoretical interest, but may have practical applications. As an example it is demonstrated that Sommerfeld's approximate solution gives Ingard's (1951) approximate solution which is valid for arbitrary source and receiver heights. [DOI: 10.1121/1.1904303]

PACS numbers: 43.20.Bi, 43.20.El, 43.28.En, 43.50.Vt [RR] Pages: 3389–3392

I. INTRODUCTION

The most important and best reference work on reflection of spherical waves is still, without a doubt, the 72 page paper by Sommerfeld (1909). He was strongly motivated by the then newly established wireless telegraphy of Maxwell, Hertz, Tesla, and Marconi (among others). Besides electromagnetic waves [Collin (2004)] the problem is also important for outdoor sound propagation [Attenborough (1988); Embleton (1996); Rudnick (1957)]. The importance of the problem is reflected in the amount of literature on this subject. The complete list of literature is not included here, but good selected lists are given by Brekhovskikh (1999), Salomons (2001), and others [Banos (1966); Bies (1996); Mechel (2002)]. Due to the nice physical interpretation and the numerical convenience the method of complex images [Di (1993); Keller (1979); Li (1996); Morse (1944); Rawlins (1983)] should be mentioned. The contributions of Rudnick (1947) and Ingard (1951) must also be mentioned since they have become classical. Chien and Soroka (1975, 1980) presented an approximate solution which is used frequently, and in particular in the Nord2000 [Krag (2002)] and the HARMONOISE (2002) engineering sound propagation methods.

The main contribution here is essentially given by a further rewriting of the exact solution given by Weyl (1919) and Ingard (1951). This results in an exact formula for the boundary loss as a function of only two complex dimensionless quantities ρ and τ. This is the main result, and the result stated in the abstract follows immediately from this. The first quantity ρ is related to the numerical distance already identified in the approximate solution given by Sommerfeld (1909). The second complex quantity τ is introduced here in Eq. (12) below, and is hereafter called the second numerical distance. By neglecting the weak dependence on the second numerical distance the result is the approximate solution given by Ingard (1951). As is well known, this solution is closely related to the solution given by Sommerfeld (1909).

a)Electronic mail: Gunnar.Taraldsen@sintef.no

II. FORMULATION OF THE PROBLEM

Let ∇^2 be the three-dimensional Laplace operator. The problem of a point source above an impedance plane as indicated in Fig. 1 is defined mathematically by the boundary value problem for the inhomogeneous Helmholtz equation

$$(-\nabla^2-k^2)p=4\pi\delta_{(0,0,h_s)} \quad \text{for } z>0, \tag{1}$$

$$\frac{\partial p}{\partial z}=-ik\beta p \quad \text{for } z=0, \tag{2}$$

for the complex pressure p with the restrictions $\mathrm{Re}\,k>0$, $\mathrm{Im}\,k\geqslant 0$, $\mathrm{Re}\,\beta>0$, $h_s\geqslant 0$, where k is the wave number and β is the specific normalized admittance of the impedance plane. The specific normalized impedance is given by $Z=1/\beta$. The point source is represented by the delta function $\delta_{(0,0,h_s)}$ where h_s is the source height. The task is to determine the pressure at an arbitrary receiver position given by the coordinates $(z,r)=(h_r,r)$ as indicated in Fig. 1. The important special cases where $h_s=0$, $h_r=0$, or $\mathrm{Im}\,k=0$ are defined by limits of cases where these parameters are nonzero. This ensures uniqueness of the solution also in these important special cases. It turns out that these limiting cases can be obtained by analytic continuation.

The solution of the boundary value problem (1) and (2) is conventionally expressed by

$$p=p_d+p_r, \tag{3}$$

$$p_d=\frac{e^{ikR_1}}{R_1}, \tag{4}$$

$$p_r=Q\frac{e^{ikR_2}}{R_2}, \tag{5}$$

$$Q=[R_p+(1-R_p)F], \tag{6}$$

$$R_p=\frac{\delta-\beta}{\delta+\beta}, \tag{7}$$

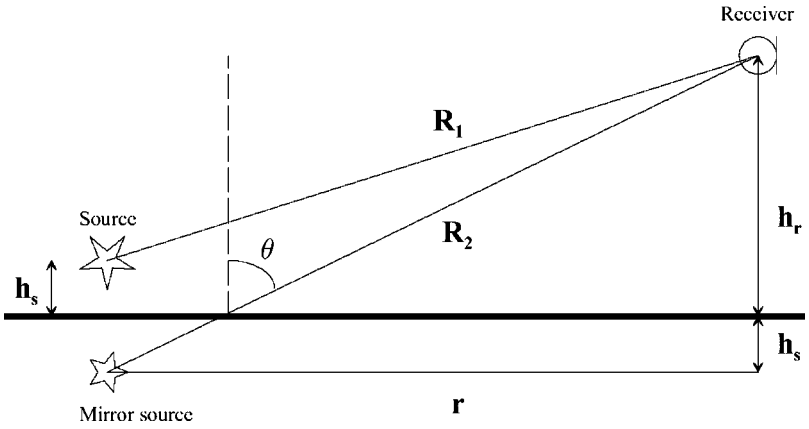


FIG. 1. Geometry of sound source and receiver above a plane.

where $\delta=\cos\theta=h/R_2$, $h=h_s+h_r$. This defines the direct wave p_d, the reflected wave p_r, the spherical reflection coefficient Q, the plane wave reflection coefficient R_p, and the boundary loss factor F. The geometrical quantities are given in Fig. 1. With this reformulation the problem of determination of p is equivalent with the determination of either Q or F.

If both source and receiver are at the ground, then the total wave solution is

$$p_0=2F\frac{e^{ikr}}{r}. \tag{8}$$

This offers an interpretation of F. Sommerfeld (1909) introduced the boundary loss factor F for this case. He proved that F for $kr\gg1$ is a function of only one complex quantity; the numerical distance. This result is by no means obvious and is perhaps the main theoretical result in Sommerfeld's paper. Sommerfeld plotted the real and imaginary part of F as a function of his numerical distance. The numerical distance is proportional with distance r, and F approaches zero for large numerical distances. Sommerfeld interpreted the numerical distance as the phase difference between the surface wave and the air wave, and observed that at grazing incidence and for hard ground the numerical distance can be small even for large distances.

III. AN OBSERVATION FROM AN EXACT SOLUTION

Weyl (1919) presented an alternative approach for the solution of the problem considered initially by Sommerfeld. The solution contained no explicit boundary wave term and the resulting asymptotic expansion is not limited to the case where both source and receiver are at the ground. Ingard (1951) observed that the impedance boundary condition gives a simplification compared to the two-medium case if Weyl's rather than Sommerfeld's approach is used. The argument given by Weyl gives Eq. (10) in Ingard's (1951) paper, and this equation simplifies further to

$$F=1-\int_C\frac{e^{-u}du}{\sqrt{W}}, \tag{9}$$

$$W=1-\frac{u}{\rho}+\left(\frac{u}{2\tau}\right)^2, \tag{10}$$

$$\rho=\frac{i}{2}kR_2\frac{(\beta+\delta)^2}{1+\beta\delta}, \tag{11}$$

$$\tau=\frac{i}{2}kR_2(\beta+\delta), \tag{12}$$

where the path of integration C is from $u=0$ to $u=\infty$ on the real axis for typical cases of interest. For the general case the path C depends on the numerical distance ρ and the second numerical distance τ. The $\sqrt{W}$ equals 1 at $u=0$ and is determined otherwise by analytic continuation along the path of integration. An alternative solution is to keep the integration path on the real axis and add a boundary wave term corresponding to integration along a branch cut depending on the location of the square-root singularity [Thomasson (1976)].

Ingard pointed out the symmetry in the β and δ dependence of F. This result is not clear from the original exact solution given by Sommerfeld. The normal incidence case $\delta=1$ (or $\beta=1$) gives $\rho=\tau$ and Ingard's exact result

$$F=1+2\rho\exp(-2\rho)E_1(-2\rho),\quad \delta=1\ \text{ or }\ \beta=1. \tag{13}$$

Here E_1 is the exponential integral function [Abramowitz (1972)]. This result does not follow immediately from Sommerfeld's solution due to a singularity for $r=0$ in the boundary wave term.

The dependence of F on τ in Eq. (9) is weak, and by neglecting this term the result is the asymptotic solution of Ingard in terms of the complementary error function [Abramowitz (1972)],

$$F\approx1+i\sqrt{\pi}\rho^{1/2}e^{-\rho}\operatorname{erfc}(-i\rho^{1/2}),\quad kR_2\gg1. \tag{14}$$

By adding and subtracting the term $(u/2\tau)^2$ in the expression for W it can be argued that Eq. (13) also can be taken as a general approximate solution for the case $kR_2\gg1$.

The approximation (14) was also derived by Lawhead and Rudnick (1951), but with ρ replaced by

$$w=|w|e^{ib}=\frac{i}{2}kR_2(\beta+\delta)^2. \tag{15}$$

Lawhead and Rudnick refer to b as the phase and the absolute value $|w|$ as the numerical distance. This is consistent with Sommerfeld's definition. The approximation (14) has been derived by different methods and in particular by Chien and Soroka (1975, 1980). This particular result by Chien and

Soroka is identical with the result derived by Lawhead and Rudnick, but Chien and Soroka choose to refer to

$$p_e = \sqrt{w} \tag{16}$$

as the numerical distance. These results are consistent with the result given above by Eqs. (11) and (14) since Chien and Soroka assume $\beta \ll 1$ and $\delta \ll 1$.

The main observation now is however that the exact boundary loss factor F depends on only two complex arguments ρ and τ. This leads to

Theorem 1: *Let $p_0(r,k,\beta)$ be the total wave field solution for the case where both source and receiver are at the boundary. The boundary loss factor F for arbitrary source and receiver heights is then given by*

$$F = \left[\frac{e^{ik_\delta R_2}}{R_2}\right]^{-1} \frac{p_0(R_2, k_\delta, \beta_\delta)}{2}, \tag{17}$$

$$k_\delta = k(1+\delta\beta), \tag{18}$$

$$\beta_\delta = \beta \frac{1+\delta/\beta}{1+\delta\beta}. \tag{19}$$

Proof: Equations (8) and (9) together with $\rho = (i/2)k_\delta R_2(\beta_\delta)^2 = (i/2)kR_2(\beta+\delta)^2/(1+\beta\delta)$ and $\tau = (i/2)k_\delta R_2 \beta_\delta = (i/2)kR_2(\beta+\delta)$ give the claim. □

It can be remarked that the conditions $\mathrm{Re}\, k_\delta \geqslant 0$, $\mathrm{Im}\, k_\delta \geqslant 0$, $\mathrm{Re}\, \beta_\delta \geqslant 0$ are typically fulfilled for sufficiently small δ, but may be violated in the general case. In this case the solution p_0 in Eq. (17) is defined by analytic continuation along the paths in the complex k and β planes given by the parameterization by δ in Eqs. (18) and (19), respectively. This will be explained by two examples.

Theorem 1 gives an alternative route for the solution of the problem with a point source over an impedance plane. In the first step it is only necessary to consider the grazing incidence case $\delta = \cos\theta = 0$. The exact result in this case can be written as a complex line integral, and Eq. (9) represents only one of many possible formulations. The analytic continuation required for the general case can typically be done simply by deformation of the integration path in the complex line integral. A particular example is given by the original Sommerfeld integral solution, and the result is then that the apparent singularity in the surface wave term for $r=0$ is transformed into a singularity for $R_2 = 0$.

This alternative route can also be used on any of the many available approximate solutions. In this case the analytic continuation is typically trivial since the functions involved are analytic in the required domains. A particular example is given by Sommerfeld's original approximate solution, and this will be considered next.

Sommerfeld (1909) considered the problem of a pulsating vertical electric dipole above a plane ground. The air and the ground are characterized by the dielectric constant, the magnetic permeability (μ), and the conductivity. The electric and magnetic fields are determined by the Hertz vector potential which is vertically directed for this problem. The corresponding z-component Π fulfills the inhomogeneous Helmholtz equation (1) with complex wave numbers k_1 and k_2 for the air and ground domains, respectively. The boundary condition (2) is replaced by continuity of Π and $\chi^{-2}\partial\Pi/\partial z$ at the boundary [Sommerfeld (1909)]. A corresponding acoustic problem is formally obtained by a replacement of the quantity $\chi^2 = k^2/\mu$ by a complex mass density. In summary the ground can be characterized by a complex wave number and a complex mass density in the acoustic problem, and characterized by a complex wave number k_2 and a complex χ_2 in the radio wave problem. Air is characterized by k_1 and χ_1.

Sommerfeld (1909) [Eqs. (47) and (55)] obtained two solutions for $kr \gg 1$ and $h=0$ for the case of large and small numerical distance. The two Sommerfeld solutions coincide (!) and are given by Eq. (14) except for a factor 2 due to a different boundary condition and a replacement of $\sqrt{\rho}$ by α, where [Sommerfeld (1909)] [Eq. (41)]

$$\alpha^2 = \frac{i}{2} k_1 r \frac{\chi_1^4}{\chi_2^4}\left(\frac{k_2^2}{k_1^2} - 1\right). \tag{20}$$

The two-medium case treated by Sommerfeld contains the boundary condition case in the limit $k_2 \gg k_1$, with the identification $\beta = (k_2\chi_1^2)/(k_1\chi_2^2)$. It can be concluded that Sommerfeld's approximation gives Eq. (14) for the case $h=0$.

Theorem 1 can be applied to Sommerfeld's result in the form of Eq. (14) for the case $h=0$. This gives an approximation for $h>0$, and the result coincides with the asymptotic result (14) of Ingard. The required $\rho^{1/2}$ in formula (14) is given by

$$\rho^{1/2} = \frac{1+i}{2}\sqrt{kR_2}(\beta+\delta)/\sqrt{1+\beta\delta}, \tag{21}$$

where $\sqrt{\cdot}$ denotes the principal value of the square-root. This is the required analytic continuation from the case $h=\delta=0$ as described after Theorem 1. The proof follows by observing that the involved functions in formula (14) are entire functions and $\sqrt{\cdot}$ is defined with the cut along the negative real axis. Incidentally formula (21) also generalizes the result of the discussion of Stinson (1995) to include complex k and the numerical distance given by Ingard.

IV. CONCLUDING REMARKS

It is remarkable that the exact boundary loss is a function of only two complex quantities ρ and τ. The resulting Theorem 1 gives the possibility of obtaining alternative approximations from existing approximations [Attenborough (1980)]. Theorem 1 can also be used to transform exact results [Enflo (1987) and Sommerfeld (1909)] which are singular at $r=0$ into exact results which are singular only at $R_2=0$. A natural task would be to obtain an approximate solution which improves Eq. (14) *and* is a function of only (ρ,τ). One possibility is to consider F in Eq. (9) as a function of $x = 1/\tau^2$ and a Taylor expansion around $x=0$.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge remarks from and discussions with Ulf Kristiansen, Peter Svensson, Tor Erik Vigran (NTNU), and Simon Neil Chandler-Wilde (The Univer-

sity of Reading). Helpful comments from two reviewers on an earlier version of the manuscript are acknowledged.

Abramowitz, M. and Stegun, I. A., editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. (Dover, New York, 1972).

Attenborough, K., ''Review of ground effects on outdoor sound propagation from continuous broadband sources,'' Appl. Acoust. **24**, 289–319 (1988).

Attenborough, K., Hayek, S. I., and Lawther, J. M., ''Propagation of sound above a porous half-space,'' J. Acoust. Soc. Am. **68**, 1493–1501 (1980).

Banos, A., *Dipole Radiation in the Presence of a Conducting Half-Space* (Pergamon, Oxford, 1966).

Bies, D. A. and Hansen, C. H., *Engineering Noise Control*, 2nd ed. (E & FN SPON, New York, 1996).

Brekhovskikh, L. M. and Godin, O. A., *Acoustics of Layered Media I–II*, 2nd ed. (Springer, Berlin, 1999).

Chien, C. F. and Soroka, W. W., ''Sound-propagation along an impedance plane,'' J. Sound Vib. **43**, 9–20 (1975).

Chien, C. F. and Soroka, W. W., ''A note on the calculation of sound-propagation along an impedance surface,'' J. Sound Vib. **69**, 340–343 (1980).

Collin, R. E., ''Hertzian dipole radiating over a lossy earth or sea: Some early and late 20th-century controversies,'' IEEE Antennas Propag. Mag. **46**, 64–79 (2004).

Di, X. and Gilbert, K. E., ''An exact Laplace transform formulation for a point source above a ground surface,'' J. Acoust. Soc. Am. **93**, 714–720 (1993).

Embleton, T. F. W., ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).

Enflo, B. O. and Enflo, P. H., ''Sound wave propagation from a point source over a homogeneous surface and over a surface with an impedance discontinuity,'' J. Acoust. Soc. Am. **82**, 2123–2134 (1987).

HARMONOISE, Harmonized accurate and reliable methods for the EU directive on the assessment and management of environmental noise. *Project funded by the European Commission: IST-2000-28419*, 2002.

Ingard, U., ''On the reflection of a spherical sound wave from an infinite plane,'' J. Acoust. Soc. Am. **23**, 329–335 (1951).

Keller, J. B., ''Progress and prospects in the theory of linear wave propagation,'' SIAM Rev. **21**, 229–245 (1979).

Krag, J., Plovsing, B., Storeheier, S. Å., Taraldsen, G., and Jonasson, H. G., ''Nordic Environmental Noise Prediction Methods. Nord2000 Summary Report. General Nordic Sound Propagation Model and Applications in Source-Related Prediction Methods,'' DELTA Acoustics & Vibration Report No. 1719/01 (**2002**).

Lawhead, R. B. and Rudnick, I., ''Acoustic Wave Propagation Along a Constant Normal Impedance Boundary,'' J. Acoust. Soc. Am. **23**, 546–549 (1951).

Li, Y. L. and White, M. J., ''Near-field computation for sound propagation above ground—Using complex image theory,'' J. Acoust. Soc. Am. **99**, 755–760 (1996).

Mechel, F. P., *Formulas of Acoustics* (Springer, Berlin, 2002).

Morse, P. M. and Bolt, R. H., ''Sound waves in rooms,'' Rev. Mod. Phys. **16**, 69–150 (1944) (The line source solution is on p. 143).

Rawlins, A. D., ''Note on a result of Morse and Bolt,'' J. Sound Vib. **89**, 365–369 (1983).

Rudnick, I., ''The propagation of an acoustic wave along a boundary,'' J. Acoust. Soc. Am. **19**, 348–356 (1947).

Rudnick, I., ''Propagation of sound in the open air,'' in *Handbook of Noise Control*, edited by C. M. Harris (McGraw–Hill, New York, 1957), Chap. 3, pp. 3.1–3.17.

Salomons, E. M., *Computational Atmospheric Acoustics* (Kluwer, Dordrecht, 2001).

Sommerfeld, A., ''Über die ausbreitung der wellen in der drahtlosen telegraphie'' (''On the propagation of waves in wireless telegraphy''), Ann. Phys. **28**, 665–737 (1909).

Stinson, M. R., ''A note on the use of an approximate formula to predict sound fields above an impedance plane due to a point source,'' J. Acoust. Soc. Am. **98**, 1810–1812 (1995).

Thomasson, S.-I., ''Reflection of waves from a point source by an impedance boundary,'' J. Acoust. Soc. Am. **59**, 780–785 (1976).

Weyl, H., ''Ausbreitung elektromagnetischer wellen über einem ebenen leiter'' (''Propagation of electromagnetic waves over a conducting plane''), Ann. Phys. **60**, 481–500 (1919).

Moving average fields, macro-scale response measures, and homogenizing micro-scale variation

John J. McCoy[a)]
The Catholic University of America, Washington, DC 20064

(Received 11 August 2004; revised 1 March 2005; accepted 25 March 2005)

There are two critical issues when deriving a macro-scale prediction model starting from a more complete, underlying model. The first is the precise relationship of the fields predicted by the more complete model and the fields predicted by the macro-scale model. The second is the manner of solving a closure problem that is invariably encountered in all such derivations. The understanding that moving averages of the fields predicted by the more complete model are the fields predicted by the macro-scale model is challenged on the grounds that accomplishing a moving average does not eliminate micro-scale variation, it only appears to do so in one representation of the moving average field. The solution of a closure problem by assumption is challenged on the grounds that the most common assumptions are demonstrably invalid, even while leading to prediction models that can provide reasonable estimates of the macro-scale response in some scenarios. In presenting the challenges, it is further shown how a multiresolution analysis by an orthogonal wavelet system provides a framework for both precisely defining macro-scale response fields, i.e., fields from which all micro-scale variation has been eliminated, and presenting a formally exact solution for a precisely described closure problem.
[DOI: 10.1121/1.1914080]

PACS numbers: 43.20.Bi, 43.20.Gp [PEB] Pages: 3393–3401

I. INTRODUCTION

The concept of a ''moving average field'' (MAF) appears prominently in a number of derivations of theories for fluid filled porous solids, of which the Biot theory[1–3] is the most often cited. Setting aside for a moment the specifics of any particular derivation, the role assigned to MAFs is to represent macro-scale measures of the response fields that would be predicted using a more complete formulation that applies to all length scales. Biot's theory, for example, is expressed in solid phase displacement, stress and strain fields and fluid phase velocity, pressure and strain rate fields, all of which vary on length scales that are far larger than the micro-scales on which the pore structure can be observed. Assuming a more complete formulation underlies Biot's theory, this more complete formulation would be expressed in response fields that vary on both macro-scales and micro-scales. The first task in deriving Biot's theory based on this more complete formulation, then, is to relate the response fields in terms of which the two formulations are separately expressed. The MAF concept accepts the macro-scale fields predicted using the Biot theory are moving averages of the macro-/micro-scale fields predicted using the more complete formulation.

The above suggests a question: Is a MAF an appropriate macro-scale response field? One seemingly obvious requirement for an appropriate macro-scale response field is that this ought include only macro-scale variation, and not any variation that is observed on micro-scales. The point to emphasize is that the act of accomplishing a moving average ought not simply reduce the strength of micro-scale variation vis-a-vis that of macro-scale variation in one representation of the field, a reduction that might be reversed by an appropriate signal processing. Rather the averaging ought to effectively eliminate micro-scale variation entirely, thereby precluding its recovery by any signal processing.

A MAF is presented in this paper as one extreme of what are termed ''fields of local averages'' (FLAs). A FLA is constructed in two steps. The first step is to collect a sufficiently dense set of discrete local averages of the underlying field; the second step is to construct a continuous field from the collected averages. Distinguishing different FLAs are differences in the precise meaning of ''sufficiently dense'' and in the manner of ''constructing a continuous field.'' Thus, a MAF is a limiting FLA constructed from local averages collected at vanishing distance between neighboring averaging intervals. The opposite limiting FLA is constructed from local averages collected for neighboring averaging intervals that are separated by a distance equal to the size of the interval. The MAF is shown to only appear not to include micro-scale variation in one representation; the averaging incorporated therein simply changes the relative strengths of the micro-scale and macro-scale contributions in this one representation. The other extreme FLA constructed from the least densely populated set of all sufficiently dense locations, on the other hand, is shown to not include micro-scale variation in all representations; the averaging incorporated therein effectively reduces the strength of micro-scale variation to zero.

In demonstrating the latter conclusion the FLA is identified with one component of a wavelet-based resolution of the underlying field, the so-called ''smooth component.''[4] The FLA is identified with the output of a low-pass spatial filter for the underlying response field as input.[5]

[a)]Electronic mail: mccoy@cua.edu

The manner of deriving ''homogenized field equations'' that govern macro-scale response fields is next considered. The critical step in any homogenization is addressing a closure problem, occasioned by an interchange of operations that do not commute. The noncommuting operations encountered are the averaging that determines macro-scale response fields and the multiplying of the underlying response fields with fields that describe the environment microstructure. For a porous media, the environment fields are masking functions that distinguish the fluid and solid phases when these are observed on the micro-scales. As first encountered, the closure ''problem'' is not a mathematical problem at all, but rather is an obstacle preventing completion of the derivation process. Two approaches, then, can be put forth for addressing its solution. The first is to simply accept the obstacle is in effect a problem in the mathematical sense, and intuit its solution, in the form of a ''closure assumption'' that expresses one's understanding of the underlying physics. The theories for fluid filled porous solids expressed in MAFs, e.g., those presented in Refs. 6–8, accept a closure assumption. A more rational approach for addressing the closure problem is to identify the mathematical problem that is the genesis of the obstacle, which can then to be solved to overcome it. The wavelet-based resolution of the complete response fields as complementary partial response fields, the ''smooth'' component identified as a FLA and the ''detail'' component, also provides a framework for partitioning the more complete formulation as complementary partial formulations expressed in the complementary partial response fields. The mathematical problem that underlies the closure problem lies within the partitioned form of the more complete problem, as shown in a later section.[9]

While the path for achieving Biot's theory as a wavelet-based homogenization can be clearly seen in the formal steps presented, the detailed steps for accomplishing this goal is left to a future research agenda for two reasons. First, the objective of the present study, as already stated, is to emphasize two issues of significance for all homogenizations, i.e., all derivations of macro-scale prediction models starting from more complete models that apply to all length scales. Another derivation of Biot's theory would not advance this objective. Second, accomplishing the detail steps to arrive at Biot's theory is recognized as a far from trivial task. To date, reported wavelet-based homogenizations[10–15] have been limited to one spatial dimension and to a more complete formulation that is described by the same field equations applying at all points in physical space. The microstructure for these scenarios is described by micro-scale variation in the environment property fields that enter the governing field equations. While limited application of a multidimensional wavelet framework has been reported,[16] its use as a basis for homogenizing field equations has not been accomplished. Then, the homogenization of formulations described by different sets of field equations applying to different spatial domains when observed on the micro-scale, e.g., the more complete formulation that underlies Biot's theory, raises an issue of the distinction between geometric and topologic length scales that is beyond the scope of this paper.

For completeness, two alternative methods for deriving macro-scale formulations are briefly considered in the light of the present study. ''Two-scale asymptotic derivations''[17,18] accommodate macro-/micro-scale variation by associating with each response field, a postulated field that is defined on a space described by two spatial coordinates, one for tracking macro-scale variation and one for tracking micro-scale variation. In providing a precisely defined resolution of a response field in ''smooth'' and ''detail'' components, the wavelet framework accomplishes by algorithmic steps what the two-scale method posits, and does so in a context that is not inherently asymptotic. The significance of the closure problem in the final result of a two-scale asymptotic derivation of Biot's equations is emphasized. Then, ''statistical smoothing derivations''[19–22] defer explicit reference to the length scales on which the response field can be observed, focusing instead on a postulated ensemble of environments for which the more complete formulation applies. In accomplishing a statistical smoothing, each randomly varying response field is ''resolved'' as a mean, ensemble-averaged, field plus a stochastic fluctuating field. Except for this different understanding of the nature of the response field resolution, the steps for accomplishing a statistical smoothing and a wavelet-based homogenization are formally identical.

II. DEFINING A FIELD OF LOCALLY AVERAGED VALUES

The motivation for introducing a FLA is recognition that any observation, or measurement, uses a physical device of finite size, which outputs a spatial average across the region spanned by the device. The act of measuring is represented by a formula,

$$U_i = \frac{1}{V}\int W(x-x_i;l_{\mathrm{dev}})u(x)dx, \tag{1}$$

where $u(x)$ is the underlying field, $W(x-x_i;l_{\mathrm{dev}})$ is a normalized window function that represents a device of size l_{dev} located at x_i, and U_i is the device output. For a device that ''clips'' the field in extracting the segment that spans the region covered and weights all points within the region uniformly, $W(x-x_i;l_{\mathrm{dev}})$ equals unity or zero for points within or without the region. The measure V is one of volume, area, or length depending on the dimension of the region covered.

A FLA is constructed by extrapolating a set of discrete local averages collected at a sufficiently dense set of measurement locations, to obtain a field. Ambiguity arises in defining sufficiently dense and in prescribing an extrapolation process. ''Sufficiently dense'' implies an upper limit on the distance separating neighboring measurement locations with rational choices delimited by two extremes. One extreme looks to the measuring device; the set of U_i values for measurement locations separated by a distance equal to l_{dev} is the most sparsely populated of all sets that warrant characterization as sufficiently dense. For this choice, all of physical space is sampled by the set of measurement locations with minimum overlap of the regions spanned for neighboring locations. A FLA constructed from this most sparsely populated set of U_i values is denoted by $U(x;l_{\mathrm{dev}})$, where l_{dev}, here, represents the distance separating neighbor-

ing measurements. For the other extreme, one looks to the underlying field and the ''smallest scale'' for observing its variation.[23] Let l_{field} denote the smallest scale for observing variation in $u(x)$ and let $U(x;l_{\text{field}})$ denote a FLA constructed from a set of local averages collected at neighboring locations that are separated by distance equal to a fraction of l_{field}. The U_i values that determine $U(x;l_{\text{field}})$ comprise a far more dense set than the most sparsely populated set that determines $U(x;l_{\text{dev}})$.

That far more measurements are incorporated in $U(x;l_{\text{field}})$ suggests the ''information content'' in $U(x;l_{\text{field}})$ is greater than in $U(x;l_{\text{dev}})$, even though the two FLAs are based on measurements using the same physical device. Intuitively, the additional information is variation in the underlying $u(x)$ field, which can be observed on scales smaller than l_{dev}. Consider: Any single U_i is one measure of the uncountable number of measures required to describe the segment of the underlying $u(x)$ field that is spanned by the device in outputting U_i. Further, the nature of the information in the segment of $u(x)$, which is not measured by a single U_i is clear; it is variation that can be observed on scales smaller than l_{dev}. Now consider a set of averages, each of which is a measure of a *different segment* of the same underlying field. At issue is the information content in the set of averages, and if any information of the variation in the underlying field on scales smaller than l_{dev} is contained therein. It is clear that if the segments of $u(x)$ do not overlap the measured averages are independent and, therefore, the information in the set is no greater than the sum of the information in its members, in isolation. This is the case for $U(x;l_{\text{dev}})$. It is equally clear, on the other hand, that if there is overlap in the segments the measured averages are not independent and, therefore, there is information in the set that is greater than the sum of the information in its members, in isolation. This is the case for the set of U_i from which $U(x;l_{\text{field}})$ is constructed. Moreover, it seems clear that the added information in this latter set are measures of the variation that can be observed on scales smaller than l_{dev}.

The second step in the construction of a FLA is the precise process for extrapolating the set of measured averages. An extrapolation by ''splining,'' a process that groups the data set in neighboring subsets from which smooth, local subfields are formed and subsequently joined to obtain a single global field, introduces its own smoothing. This obviously could eliminate small-scale variation but not in a manner that can be incorporated in the derivation of macro-scale field equations. Alternatively, an extrapolation as a ''synthesis'' is succinctly expressed by

$$U(x;l)=\sum_i U_i\phi(x-x_i;l), \tag{2}$$

where the $\phi(x-x_i;l)$ denote a set of shifted windows, functions that fall rapidly to zero outside a neighborhood of the center location, x_i. The size of the neighborhood is denoted by l, understood to equal the nearest-neighbor separation. An extrapolation as a synthesis obviously assigns values to the extrapolated field in the neighborhood of a location at which an average is measured ''exclusively'' by the value of the measured average, i.e., independent of the values of averages not in the neighborhood.

For $U(x;l_{\text{dev}})$, the size of the window that describes the averaging is equal to the size of window that describes the synthesis. It would be a small step, then, to equate the two window functions, to obtain a FLA that is described as ''wavelet-motivated'' in a subsequent subsection. For $U(x;l_{\text{field}})$, the size of the window for synthesizing the measured values equals the smallest scale for observing variation in the underlying $u(x)$ field. It would be a small step, then, to approximate this FLA by one that would be constructed from averages collected for vanishing nearest-neighbor separation, synthesized using a delta-function window.

A. A moving average field

Consider a FLA based on U_i measurements accomplished at infinitesimal nearest-neighbor separation. Mathematically, this FLA is the small l_{field} limit of $U(x;l_{\text{field}})$. Alternatively, the FLA has a single-step representation described by Eq. (1) with the discrete variables, x_i, replaced by the continuous variable, x. This alternative representation identifies the FLA as a MAF.

In discussing the MAF in detail, it is convenient to limit the measuring device to one that spans a line segment and weights all points covered uniformly. A line gauge for sampling a strain field would be a representative device. The output of a moving gauge of length L is described by

$$U(x)=\frac{1}{L}\int_{x-L/2}^{x+L/2}u(x')dx'. \tag{3}$$

Upon differentiating one obtains still another formula,

$$\frac{dU(x)}{dx}=\frac{1}{L}(u(x+L/2)-u(x-L/2)). \tag{4}$$

Referring to Eq. (4), a field, $u(x)$, that has micro-scale variation measured relative to L, leads to a right-hand side that has micro-scale variation. Clearly, then, $U(x)$, which upon differentiation equals the right-hand side, must be judged to incorporate micro-scale variation in $u(x)$. Indeed, subject to an assumption that one has auxiliary information of $u(x)$ across any fixed line segment of length L, this formula would allow complete recovery of $u(x)$ at all locations, x, based on measurements incorporated in $U(x)$.

The functions for which one is interested in constructing a MAF are not arbitrary but belong to a set that describe the response fields for certain physical experiments. It is instructive to describe this set and investigate the right-hand sides of Eqs. (3) and (4) for its members. The set can be identified with the aid of a physical device, a low-pass spatial filter. Fundamentally, a low-pass spatial filter identifies two sets of functions, those that pass the filter undistorted and those that are blocked by the filter. Intuitively, the first set is comprised of functions that vary on length scales that are significantly larger than the characteristic length of the filter. Let $u_M(x)$ represent a generic function that varies on length scales bounded from below by l_M, a minimum, macro-length that is significantly larger than L, the characteristic length of a certain low-pass filter. Further, the second set is comprised of functions that vary of length scales that are significantly

smaller than the characteristic length of the filter. Let $u_m(x)$ represent a generic function that varies of length scales bounded from above by l_m, a maximum, micro-length that is significantly smaller than L. The set of functions for which one is interested in constructing a MAF is comprised of a synthesis of one function from each of these two sets. Let $u(x)=u_M(x)+u_m(x)$ represent the generic element of this last set. Clearly, the description of the last set is consistent with the macro-/micro-scale language used for describing the physical experiments for which a simplified, macro-scale prediction model is warranted.

The filtering language settles, in one sense, the issue of the precise relation of the response fields predicted by a macro-scale model and those predicted by the more complete model that underlies it, by identifying the former as the output of low-pass filter when the latter is the input. This identification, of course, begs the question of describing the *mathematical operator* that represents the action of a low-pass filter, a description that is required if one is to "derive" a model for predicting the filter output starting from the more complete model for predicting the filter input. The filtering language is useful to the discussion of this section in restating the question of the efficacy of the concept of a MAF: Does the constructed moving average when applied to the restricted class of response fields reproduce the macro-scale component as it blocks the micro-scale component? The emphasis in the wording is the requirement that the micro-scale component be blocked, and therefore unrecoverable using any signal processing.

The restated question is addressed by investigating the right-hand sides of Eqs. (3) and (4) for elements of the restricted set. This can be accomplished in general, or in the context of an idealized example that represents the set; the interested reader might use $u(x)=A_M \sin(x/l_M)+A_m \sin(x/l_m)$. Substituting either the general form or the example into the right-hand side of Eq. (3) leads to the conclusion that $U_M(x)$, the moving average of $u_M(x)$, essentially reproduces $u_M(x)$, and that the "strength" of the variation in $U_m(x)$, the moving average of $u_m(x)$, is a small fraction, l_m/L, of the strength of the variation in $u_m(x)$. It is the sharp reduction of the strength of the variation in $u_m(x)$ in contributing to $U(x)$ that suggests the construction of a moving average represents the action of a low-pass filter. However, as is made particularly clear using the idealized example, the *variation in the much reduced $U_m(x)$ is observed on the micro-scales*.

This conclusion is further emphasized on substituting either the general form or the specific example into the right-hand side of Eq. (4). The result for the general form is written,

$$\frac{dU}{dx}=\frac{\Delta u_m}{L}+\frac{\Delta u_M}{L}=\frac{\Delta u_m+\Delta u_M}{L}. \tag{5}$$

Writing the right-hand side as $\Delta u_m/L+\Delta u_M/L$ and noting that the variation in $u_M(x)$ is observed on scales that are significantly larger than L, leads one to conclude that the second term in this combination is a good approximation of du_M/dx. Thus, would one be able to conclude this second term dominated the first term, one would arrive at a conclusion that dU/dx is a good approximation of du_M/dx, consistent with a low-pass filtering understanding. Writing the right-hand side as $(\Delta u_m+\Delta u_M)/L$, however, emphasizes that the correct comparison of the contributions of u_M and u_m in determining dU/dx is the difference in their values at the two ends of the averaging interval. Given that $u_M(x)$ is essentially a constant over the extent of the averaging interval, the "strength" of the variation in Δu_M with x, a variation that is observed on the macro-scale, is a small fraction, L/l_M, of the strength of the variation in $u_M(x)$. This is not the case for the first term, Δu_m; the strength of its variation with x, a variation that is observed on the micro-scale, remains of the same order as the strength of the variation in $u_m(x)$. In order to compare the relative strengths of the variations in Δu_M and Δu_m to one another, one obviously must introduce the relative strengths of the variations in u_M and u_m to one another. This latter comparison depends on the specific response field that u represents. For u representing a displacement field, for example, the strength of the variation in u_m is a small fraction of the strength of the variation in u_M, which subsequently leads to a conclusion that the strengths of the variations in Δu_m, observable on the micro-scales, and in Δu_M, observable on the macro-scales, are of the same order. For u representing a strain field, on the other hand, the strengths of the variation in u_m and in u_M are equal, which leads to a conclusion that the variation in Δu_m, observable on the micro-scales, dominates the variation in Δu_M, observable on the macro-scales. *For no response field of interest does the variation in Δu_M dominate the variation in Δu_m.*

The above-noted consideration simply emphasizes that a MAF is not a useful measure for isolating macro-scale variation; it only *appears* to isolate this variation in *one representation* of the MAF. Micro-scale variation continues to be present and is easily observed in other representations of the MAF. This, as demonstrated in Sec. II B, is not the case for the other extreme FLA.

In concluding this section, one might note that introducing the concept of a low-pass filter and then describing a mathematical operator to represent it actually inverts a more rational approach of first describing the mathematical operator and then showing that this represents a low-pass filter. The reason why the latter approach is the more rational is that there are numerous mathematical operators that represent numerous kinds of low-pass filters, not all of which present a framework for deriving macro-scale prediction models. The process was inverted in this section only to emphasize a point; the mathematical operator that outputs a MAF does not represent a low-pass filter. In Sec. II B, the correct order is re-established with the introduction of a wavelet-based representation of a field, which both describes an operator that can be shown to represent a low-pass filter and presents a framework for deriving a model for predicting the low-pass filter output.

B. A wavelet-motivated FLA

The second FLA to be investigated in greater detail is based on the most sparsely populated, sufficiently dense set of measurements and a synthesizing window that is identical

to the ''resolving window,'' i.e., the window representing the averaging device. With a slight change is notation, Eq. (1) can be rewritten as

$$U_i = \langle W_i(x), u(x) \rangle = \frac{1}{V} \int W_i(x) u(x) dx, \tag{6}$$

where the angular brackets interpret the averaging integral as a scalar product. Geometrically, this interpretation identifies the set of local averages as projections of the underlying $u(x)$ onto a set of directions in the function space within which $u(x)$ is to be found; each direction identified by a shifted replica of the averaging window. Next, equating $\phi(x-x_i)$ and $W(x-x_i)$, the synthesis formula, Eq. (2), can be rewritten as

$$U(x;l_{\text{dev}}) = \sum_i \langle W_i(x), u(x) \rangle W_i(x). \tag{7}$$

This, then, identifies $U(x;l_{\text{dev}})$ as a ''projection'' of $u(x)$ onto a subspace of the full space of which $u(x)$ is a member; the coordinate axes for the function subspace are described by the set of shifted windows, $W_i(x)$.

As evocative as the interpretation of Eqs. (6) and (7) may be, demonstrating the interpretation is warranted and determining the precise relationship of the subspace of which the $U(x;l_{\text{dev}})$ is a member to the full space of which $u(x)$ is a member, is a nontrivial mathematical task. Fortunately a part of the recently formulated theory of wavelets, termed a ''multiresolution analysis,'' addresses the issues in a mathematically rigorous manner.[4] This framework holds promise as a definitive tool for homogenizing field equations that govern the response of continua with complex, small-scale structure, thereby obtaining equations expressed entirely in macro-scale response measures.

Given the existence of an extensive literature on wavelets and the limited objectives of this paper, a further exposition of a multiresolution analysis is neither necessary nor warranted. A few simple elements of the theory are useful, however, for appreciating its appropriateness both for identifying macro-scale response fields and as a framework for deriving ''homogenized'' formulations governing these measures, the topic of Sec. III. Thus, the wavelet theory refers to a class of different ''representations'' of a time series or spatial field, each separate representation associated with a different wavelet system. A wavelet system is defined by two precisely related, scale-independent functions termed a ''scaling function'' and a ''mother wavelet.'' In the context of the desired application, once an arbitrary scale is introduced, the scaling function can be identified with the averaging function, $W(x)$, which defines the bases for Eqs. (6) and (7). In the context of the wavelet theory, then, $U(x;l_{\text{dev}})$ is associated with the ''smooth'' component of the underlying field $u(x)$, understood to be its ''projection'' on a *subspace of smooth functions*, i.e., functions that have no variations on length scales that are significantly smaller than an arbitrarily introduced scale. This is the basis of the filtering language introduced in Sec. II A and the identification of $U(x;l_{\text{dev}})$ with the output of a wavelet-defined low-pass filter. For the purpose of identifying macro-scale response fields, one need not delve any further into the wavelet theory.

For the purpose of deriving homogenized formulations starting with a more complete description of the physics that applies to all length scales, on the other hand, one needs the mother wavelet. The role of the mother wavelet is to define a basis for a complementary subspace to the subspace of smooth functions, in forming the function space to which $u(x)$ belongs. This complementary subspace is termed the *subspace of detail functions*; the low-pass filter described by the scaling function completely suppresses all members of this subspace. Restricting the wavelet systems to a class designated as orthogonal, the subject matter of multiresolution analysis, the subspaces of smooth and of detail functions are *orthogonal* thereby allowing the definition of a second projection operator that maps a generic $u(x)$ onto the subspace of detail functions. The result of this second projection, denoted by $u^{(d)}(x)$, is termed the ''detail'' component of the $u(x)$. This component is understood in terms of the output of a wavelet-defined high-pass filter.

The elements of the wavelet theory that provides for its usefulness as a basis for a ''homogenization'' *that is not inherently asymptotic* are: (1) the one-to-one resolution of a generic field $u(x)$ as two components $u^{(s)}(x)$ and $u^{(d)}(x)$, one of which is an appropriate macro-scale field; and (2) the definition of complementary projection operators for ''partitioning'' a more complete formulation that applies for all length scales, as coupled, partial formulations that separately apply to macro-scales and micro-scales. The use of the wavelet theory as a basis for homogenization is described further in Sec. III.

III. HOMOGENIZING GOVERNING EQUATIONS: A CLOSURE PROBLEM

The mathematical task of deriving field equations that govern macro-scale response fields starting from a more complete formulation that applies to all length scales can be described as a homogenization. A critical step in a homogenization is addressing a closure problem occasioned by two operations that do not commute, one an operation of averaging and the other a multiplication by an environmental field that is part of the more complete formulation. The closure problem is encountered in averaging one or more of the equations that are to be homogenized. As an illustration, consider a fairly typical example of a constitutive equation relating stress and strain fields,

$$\tau(x) = C(x)\epsilon(x). \tag{8}$$

In this equation $C(x)$ is the environmental field, in this case a heterogeneous property field. Letting braces denote an average, the average of Eq. (8) is written,

$$\{\tau(x)\} = \{C(x)\epsilon(x)\}. \tag{9}$$

An interchange of the order of averaging and multiplying would obtain a right-hand side equal to $\{C(x)\}\{\epsilon(x)\}$, an equation relating the averaged stress and averaged strain fields. The interchange, however, would only be valid for commuting operators, which is not the case here. Accepting the operations do not commute one can always express what is in fact a nonequation by an equation, on introducing an additional term. One writes

$$\{\tau(x)\}=\{C(x)\}\{\epsilon(x)\}+\sigma(x), \tag{10}$$

an equation, which as derived, serves to "define" $\sigma(x)$. Consider, however, the possibility of another definition of σ, one that determines it as a mapping of $\{\epsilon(x)\}$ and, perhaps, other response fields that enter the homogenized formulation. Introducing this mapping, or rule, into Eq. (10) would result in an equation expressed entirely in response fields that enter the homogenized formulation. It is in this sense that the resulting equation is said to be "closed." Determining the mapping is the heart of the closure problem.

Underlying the concept of a closure problem is an implicit assumption; there exists an underlying problem that can be used to define the mapping that determines $\sigma(x)$. A rational approach to the closure problem, then, would be to first formulate this underlying problem, which could be subsequently solved to determine the mapping. An ad hoc approach to the closure problem would simply accept the underlying problem exists, without identifying it, and intuit its solution. The ad hoc approach is to solve the closure problem by assumption.

The ad hoc and rational approaches to the closure problem are considered separately in Secs. III A and III B.

A. Accomplishing "closure" by assumption

In studies formulating macro-scale equations for fluid filled porous solids, which are expressed in MAFs defined on the response fields, the closure problem is encountered in relating the MAF defined on the gradient of a response field, accomplished across those segments of the averaging interval within which the response field is defined, to the gradient of the MAF defined on the response field itself. The environmental measure in this case is a masking function that describes the segments spanned by either the solid frame or the fluid filled pore space.

Consider the formula that expresses a MAF accomplished across those segments of an averaging interval within which a response field measure $u(x)$ is defined,

$$U(x)=\frac{1}{L}\int_{x-L/2}^{x+L/2} m(x')u(x')dx'. \tag{11}$$

The masking function $m(x)$ equals unity for points within the region for which $u(x)$ is defined, and zero otherwise. The closure problem is encountered in relating the gradient of $U(x)$ to the MAF defined on du/dx.[24] The genesis of the problem, here, is the operations of differentiating within and integrating across subregions of the line segment $x-L/2<x'<x+L/2$, do not commute. Thus,

$$\left\{\frac{du(x)}{dx}\right\}\neq\frac{dU(x)}{dx}.$$

The inequality can be expressed as an equality on introducing a term, $G(x)$, writing

$$\left\{\frac{du(x)}{dx}\right\}=\frac{dU(x)}{dx}+G(x). \tag{12}$$

Closure is accomplished by a mapping that determines $G(x)$ in terms of $U(x)$ and the MAFs defined on the other response fields that appear in the homogenized formulation.

The first approach to obtaining the mapping is to assume its form. This is the approach followed in achieving theories of poroelasticity expressed in moving average response fields. The assumption is not introduced at this stage of the derivation, however, but is delayed until after a manipulation to which the term "averaging theorem" is applied. The averaging theorem was first proposed in the early 1960s by Slattery[25] and is featured prominently in many studies by Whitaker,[6] Pride *et al.*,[7] Berryman and Pride,[8] and others. While the averaging theorem is presented in the referenced papers in the context of multidimensional averaging regions, the manipulations incorporated are actually more transparent with a restriction to a single dimension. Thus, differentiating Eq. (11) gives the following expression for the derivative of the MAF defined on $u(x)$:

$$\frac{dU(x)}{dx}=\frac{1}{L}(m(x+L/2)u(x+L/2) - m(x-L/2)u(x-L/2)). \tag{13}$$

Next, the MAF defined on the gradient of $u(x)$ is written,

$$\left\{\frac{du(x)}{dx}\right\}=\frac{1}{L}\int_{x-L/2}^{x+L/2} m(x')\left(\frac{du(x')}{dx'}\right)dx', \tag{14}$$

which upon integrating by parts gives

$$\left\{\frac{du(x)}{dx}\right\}=\frac{1}{L}(m(x+L/2)u(x+L/2)-m(x-L/2)u(x-L/2))-\frac{1}{L}\int_{x-L/2}^{x+L/2}\left(\frac{dm(x')}{dx'}\right)u(x')dx'. \tag{15}$$

Identifying the first term on the right-hand side of Eq. (15) with

$$\frac{dU(x)}{dx};$$

see Eq. (13); reduces the equation to Eq. (12) with $G(x)$ given by

$$G(x)=\frac{1}{L}\int_{x-L/2}^{x+L/2}\left(\frac{dm(x')}{dx'}\right)u(x')dx'. \tag{16}$$

Introducing a representation of the derivative field for the masking function, as a sum of delta function distributions, one can reduce Eq. (16) by writing

$$G(x)=\frac{1}{L}\sum_i (\pm)_i u(x_i), \tag{17}$$

where the x_i delimit those segments that are in the interval $(x-L/2,x+L/2)$, within which $u(x)$ and du/dx are defined. The sign assigned to $u(x_i)$ alternate, as determined by ones entering or leaving the segment when moving in the direction of increasing x.

It is at this point that a rule is postulated for determining $G(x)$ in terms of $U(x)$ and the MAFs defined on the other response fields that appear in the homogenized formulation. The existence of a formula that determines $G(x)$ in terms of the more fundamental $u(x)$ is obviously seen as a source of

physical insight for intuiting the closure assumption. Thus, for example, the formula demonstrates the value of $G(x)$ at a generic location, x, depends only on $u(x)$ values at locations in the *interior* of the averaging region identified by x; the mapping that determines $G(x)$ in terms of discrete u_i values is in this sense ''local.'' This is an interesting conclusion that might suggest, but hardly demonstrates, that the mapping that determines $G(x)$ in terms of $U(x)$ and the other MAFs is local in the same sense. The closure assumptions that result in Biot-type equations accept local rules.

Leaving aside the issue of the implication, or lack thereof, of Eq. (17) for postulating a closure rule for determining $G(x)$, the author confesses to being troubled by an apparent inconsistency with the interpretation of the averaging theorem *as it is proved* and *as it is applied*. Consider that the theorem expresses an equation of ''operators,'' and not functions per se. That is, Eq. (12) equates the results of the actions of certain combinations of operators on *each and every member of a function space*. The inconsistency is the description of the function space. Thus, in proving the theorem the operators involved both act upon and result in functions that have both macro-scale and micro-scale variation. For example, the operator dU/dx is understood when proving the theorem, to be a combination of two operators acting in sequence; the first is the integral operation described by the right-hand side of Eq. (11); the second is the differentiation that results in the right-hand side of Eq. (13). Clearly, the first operator acts on a function that has both macro-scale and micro-scale variation, and the second operator results in a function that has both macro-scale and micro-scale variation. Similarly, $\{du/dx\}$ and $G(x)$, the latter given by Eq. (17), are interpreted when proving the averaging theorem, as acting upon and resulting in functions that have both macro-scale and micro-scale variation. The ''proof'' that dU/dx is equal $\{du/dx\}-G(x)$ is that these combined operators when acting on a function that has both macro-scale and micro-scale variation result in the *same* function that has both macro-scale and micro-scale variation. In applying the theorem, on the other hand, Eq. (12) is used to eliminate the operator represented by $\{du/dx\}$ in favor with the operator represented by $G(x)$. Then, the operator represented by $G(x)$, one that according to Eq. (17) acts on functions that have both macro-scale and micro-scale variation, is, by the closure assumption, replaced by an operator that acts on $U(x)$, and perhaps other functions, all of which are ultimately identified as macro-scale response fields.

B. A substructuring approach to the closure problem

An understanding of $U(x;l_{\text{device}})$ as one component of a multiresolution analysis of a generic $u(x)$—the smooth component denoted by $u^{(s)}$—carries with it the identification of a complementary component—the detail component denoted by $u^{(d)}(x)$. Further, this understanding also carries with it the introduction of two precisely defined projection operators, $\mathbf{P}^{(s)}$ and $\mathbf{P}^{(d)}$ such that each when acting separately on any $u(x)$ gives $u^{(s)}(x)$ and $u^{(d)}(x)$.

The resolution of all response fields that enter the more complete formulation that applies to all length scales, and the separate actions of the complementary projection operators on the equations expressing this more complete formulation results in a ''partitioned form'' of the governing equations. To illustrate the partitioning consider a ''scattering'' formulation for a radiation field in an environment with a microstructure described by heterogeneous property fields that can vary on both the micro-scales and, perhaps, the macro-scales. The more complete model is expressed somewhat abstractly by

$$u = u_0 + \mathbf{G}_0\mathbf{Q}u, \tag{18}$$

where u is the response field, the unknown radiation field; u_0 is the corresponding field for a reference environment for which there is no micro-scale heterogeneity; $\mathbf{Q}$ is a ''local'' operator that describes the interaction of the as-yet-undetermined radiation field, u, with the micro-scale varying heterogeneity, represented by $\mathbf{Q}$; and, $\mathbf{G}_0$ is a ''propagator'' that describes the distribution of the result of the interaction of u and the microstructure throughout the reference environment. The experiment forcing is represented in this equation by the response field in the reference environment for which there is no micro-scale heterogeneity.

Introducing $u = u^{(s)} + u^{(d)}$ into Eq. (18) while projecting it using $\mathbf{P}^{(s)}$ and $\mathbf{P}^{(d)}$ results in the following pair of equations:

$$u^{(s)} = u_0 + \mathbf{P}^{(s)}\mathbf{G}_0\mathbf{Q}u^{(s)} + \mathbf{P}^{(s)}\mathbf{G}_0\mathbf{Q}u^{(d)}, \tag{19}$$

and

$$u^{(d)} = \mathbf{P}^{(d)}\mathbf{G}_0\mathbf{Q}u^{(s)} + \mathbf{P}^{(d)}\mathbf{G}_0\mathbf{Q}u^{(d)}. \tag{20}$$

The presence of the $u^{(d)}$ term in Eq. (19) is a manifestation of a closure problem. In this sense, Eq. (19) is analogous to the averaged constitutive equation used previously as motivation, i.e., Eq. (9). Now, however, Eq. (20), interpreted as an equation that determines $u^{(d)}$ in terms of $u^{(s)}$, gives a precise definition of $u^{(d)}$ that is *other than that which enforces closure*.

An assumption that is fundamental to the objective of a homogenization is the absence of any direct reference to u_0 in Eq. (20), an assumption that restricts the experimental scenarios to those for which there is no direct external micro-scale forcing. The genesis of nonzero micro-scale variation in the response fields is a result on across-scale coupling.[26]

One can formally solve Eq. (20), writing

$$u^{(d)} = (\mathbf{I} - \mathbf{P}^{(d)}\mathbf{G}_0\mathbf{Q})^{-1}\mathbf{P}^{(d)}\mathbf{G}_0\mathbf{Q}u^{(s)}, \tag{21}$$

thereby solving the closure problem. Substituting Eq. (21) into Eq. (19) gives a closed equation in $u^{(s)}$. One writes

$$u^{(s)} = u_0 + \mathbf{P}^{(s)}\mathbf{G}_0\mathbf{Q}(\mathbf{I} + (\mathbf{I} - \mathbf{P}^{(d)}\mathbf{G}_0\mathbf{Q})^{-1}\mathbf{P}^{(d)}\mathbf{G}_0\mathbf{Q})u^{(s)}. \tag{22}$$

As presented, the best that can be said of the derivation of Eq. (22) is that it is self-consistent, i.e., there is no assumption or intermediate step that is not consistent with the final equation; and is formally exact. While this characterization is far from rigorous, it is no small thing. Moreover, for one-dimensional scenarios, the multiresolution analysis framework has been shown to provide matrix representation for all the operators involved.[10–15] Thus, all the intermediate steps and the end result, i.e., Eq. (22), have algorithmic de-

scription. Specifically, Eq. (22) is represented as a matrix equation expressed in U_i, the weighted averages of the response field, $u(x)$, from which its smooth component, $u^{(s)}\times(x)$, is constructed. It is left for a future research agenda to use the multidimensional multiresolution analysis framework to provide algorithmic description of Eq. (22) for this scenario.

Also left for a future research agenda is the ''interpretation'' of the derived matrix equation that determines the discrete measures of $u^{(s)}$, identified as local weighted averages of the underlying $u(x)$ field, as ''continuum field equations'' expressed in $u^{(s)}(x)$. This is expected to be a nontrivial task as the reported studies for one-dimensional scenarios suggest. Finally left for a future research agenda are scenarios for which the more complete formulation is not described by a single set of field equations applying to all points in an environment with a microstructure described by property field heterogeneity. The more complete formulation underlying the Biot theory couples the equations of linear elastodynamics and the Navier equations, each governing the response in separate physical domains, across an interface that is observed on a micro-scale.

IV. OTHER FRAMEWORKS FOR ACHIEVING HOMOGENIZED FORMULATIONS

Accomplishing a homogenization based on the multiresolution analysis framework is not the first derivation of a macro-scale prediction model, which approaches the closure problem rationally. It is instructive to relate and compare the multiresolution analysis approach to two others that appear prominently in the literature, a two-scale asymptotic derivation and a statistical smoothing.

The two-scale asymptotic framework[17] posits a mathematics that the multiresolution framework justifies, in the limit of an unbounded macro/micro-scale gap. Thus, in resolving the variation in a response field in smooth and detail components, each varying on scales at the two extremes of a large scale gap, the introduction of different spatial coordinates for tracking the resolved variations is obviously warranted. Further, it is a small step to postulate a limit process of the scale gap increasing without bound for which the bandwidths for separately observing the micro-scale and macro-scale variations vanish at the limit point. Small though the step may be, it clearly ties the two-scale approach to the limit point; finite bandwidth effects would be a concomitant of any finite scale-gap effects; leaving an important problem, that for a large but finite scale gap. Accepting the basic postulate of the two-scale method, the next step is to ''resolve'' the underlying formulation as an infinite hierarchy ordered in powers of the micro-/macro-scale ratio. The zero-order term, once separated from the hierarchy, would then represent the end result of the derivation. Separating the zero-order term does not naturally occur, however, presenting a form of a closure problem. The details of a two-scale derivation of Biot's equations are presented in an often-cited paper by Burridge and Keller.[18] The derivation presented is instructive for demonstrating the need to address the closure problem carefully. One, among several, of the interesting conclusions of Burridge and Keller is that Biot's equations result only for environments for which the pore structure has no macro-scale variation. Otherwise, the closure accomplished requires the introduction of an additional term, the significance of which Burridge and Keller did not investigate.

In providing an algorithmic framework that justifies the seemingly arbitrary postulation of the two-scale method, the multiresolution analysis approach is not confined to the limit point of an unbounded scale gap. Thus, it represents an appropriate framework for accommodating finite bandwidths for the separate observations of macro-scale and micro-scale variation, measured relative to a finite scale gap.

Formally, the steps accomplished in the last section are identical to those identified as a ''statistical smoothing;'' see Beran and McCoy,[19,20] Keller,[21] and McCoy;[22] once one introduces the language of statistics. For a statistical smoothing derivation, the underlying formulation is accepted as applying to a random ensemble of statistically identical experiment scenarios with $u^{(s)}$ now understood to represent the mean, or ensemble-averaged, response field while $u^{(d)}$ represents the randomly fluctuation field. The projection operator, $\mathbf{P}^{(s)}$, for a statistical smoothing represents the act of ensemble averaging, while $\mathbf{P}^{(d)}$ has no other definition than $\mathbf{I}-\mathbf{P}^{(s)}$. The equation governing the mean response field, the equivalent of Eq. (22), is termed a Dyson equation in the statistical physics literature. Except for noting the formal identity of the steps in the two developments, further reference to a statistical smoothing is not warranted.

V. SUMMARY

A multiresolution analysis using discrete orthogonal wavelets is a promising tool for modeling mechanical systems that are observable on two ranges of length scales separated by a length-scale gap, a macro-/micro-scale scenario. Thus, in providing transformations from a location space to a location/scale space and back, it allows for the resolution of macro-/micro-scale variation as separate components; the resolution is accomplished by a sequence of operations that together represent a spatial filtering. Then, in providing transformations that are related as one-to-one inverses, it allows for the representation of certain invertible operators as matrices defined on the location/scale space, which have inverses. This, is turn, allows for a resolution of certain macro-/micro-scale processes as components identified as intrascale and interscale.

In this paper, the use of the multiresolution analysis for accomplishing a homogenization, i.e., deriving a macro-scale prediction model starting from a more complete macro-/micro-scale model, is described and contrasted with other approaches that appear in the literature. First, the so-called smooth component of a macro-/micro-scale variation, accomplished using a multiresolution analysis, is identified with one extreme of a class of fields of local averages (FLAs). The other extreme FLA is identified as a moving average field (MAF), accomplished as a mapping that is demonstrated to not eliminate the micro-scale component of a macro-/micro-scale variation, as would one that represents a spatial low-pass filter. Second, the resolution of macro-/micro-scale processes as interscale and intrascale compo-

nents is demonstrated to provide for a partitioning of a macro-/micro-scale prediction problem in a manner that enables a self-consistent projection of a macro-scale prediction model. The critical step in the projection is the precise identification of a mathematical problem to be solved to accomplish ''closure.'' This is in contrast to the resort to a closure ''assumption,'' which is necessary when attempting a prediction model expressed in MAFs.

Future research agenda formulated around the uses of multiresolution analysis abound for prediction modeling for multiscale mechanical systems. Thus, as previously stated, the application of multiresolution analysis to multiple dimensions remains relatively rare.[16] Then, the details of homogenizing different field equations described in complementary physical space domains, when observed on the micro-scale, which are to be integrated subject to continuity conditions that apply at the domain interface, have not been accomplished. Finally, the development of macroscale prediction models in the presence of a large-scale gap that is characteristic of macro-/micro-scale scenarios, describes only one application. Thus, for example, statistical physics prediction modeling begins but does not end with predicting the mean field. The field variance is an important measure.

[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] M. A. Biot, ''Mechanics of deformation and acoustic propagation porous media,'' J. Appl. Phys. **33**, 1482–1498 (1962).

[4] G. Kaiser, *A Friendly Guide to Wavelets* (Birkhauser, Boston, 1994).

[5] The intent is not to argue that the wavelet-based resolution is the only possible construction of a macro-scale response field; a filtering accomplished in a window-Fourier-transform described phase-space, for example, is another possible construction. For technical reason, however, the window-Fourier-transform-basis does not present a framework for deriving a macro-scale prediction model.

[6] S. A. Whitaker, ''Advances in the theory of fluid motion in porous media,'' Ind. Eng. Chem. **12**, 14–28 (1969).

[7] S. A. Pride, A. F. Gangi, and F. D. Morgan, ''Deriving the equations of motion for porous isotropic media,'' J. Acoust. Soc. Am. **92**, 3278–3290 (1992).

[8] J. G. Berryman and S. R. Pride, ''Volume averaging, effective stress rules, and inversion for microstructural response of multicomponent porous media,'' Int. J. Solids Struct. **35**, 4811–4843 (1998).

[9] The partitioning is said to be accomplished in a wavelet-defined phase space. It is only for an orthogonal wavelet system that the described phase-space basis does not over-determine the space, a condition that is necessary for the partitioned formulation to remain well posed. It is this requirement that precludes the partitioning be accomplished in a window-Fourier-transform-defined phase space.

[10] B. Z. Steinberg and J. J. McCoy, ''Toward local effective parameter theories using multiresolution decomposition,'' J. Acoust. Soc. Am. **96**, 1130–1143 (1994).

[11] B. Z. Steinberg and J. J. McCoy, ''Effective measures of nonstationary microscale stiffness variation using multiresolution decomposition,'' J. Acoust. Soc. Am. **98**, 3516–3526 (1995).

[12] B. Z. Steinberg and J. J. McCoy, ''A multiresolution analysis of scattering by a pair of local regions of complex heterogeneity,'' J. Acoust. Soc. Am. **103**, 1273–1281 (1998).

[13] B. Z. Steinberg and J. J. McCoy, ''A study of the effective properties of mass and stiffness microstructures—A multiresolution approach,'' Q. Appl. Math. **57**, 401–432 (1999).

[14] B. Z. Steinberg and J. J. McCoy, ''A multiresolution homogenization of modal analysis with application to layered media,'' Math. Comput. Simul. **50**, 393–417 (1999).

[15] B. Z. Steinberg, J. J. McCoy, and M. Mirotznik, ''A multiresolution approach to homogenization and effective modal analysis of complex boundary value problems,'' SIAM J. Appl. Math. **60**, 939–966 (2000).

[16] M. Dorobantu and B. Engquist, ''Wavelet-based numerical homogenization,'' SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **35**, 540–559 (1998).

[17] E. Sanchez-Palencia, *Non-Homogeneous Media and Vibration Theory, Lecture Notes in Physics* (Springer, New York, 1980).

[18] R. Burridge and J. B. Keller, ''Poroelasticity equations derived from microstructure,'' J. Acoust. Soc. Am. **70**, 1140–1146 (1981).

[19] M. J. Beran and J. J. McCoy, ''Mean field variations in random media,'' Q. Appl. Math. **28**, 245–258 (1970).

[20] M. J. Beran and J. J. McCoy, ''Mean field variations in a statistical sample of heterogeneous linear elastic solids,'' Int. J. Solids Struct. **6**, 1035–1054 (1970).

[21] J. B. Keller, ''Effective behavior of heterogeneous media,'' *Statistical Mechanics and Statistical Methods in Theory and Applications*, edited by R. Landman (Plenum, New York, 1977), pp. 631–644.

[22] J. J. McCoy, ''Conditionally averaged response formulations for two-phase random mixtures,'' J. Appl. Mech. **58**, 973–981 (1991). In addressing a microstructure comprised of two interpenetrating phases based on which a macro-scale prediction model having the form of a mixture theory is derived, the study reported in this reference encounters many of the same issues as in deriving Biot's theory.

[23] A precise definition of the smallest scale for variation in a field requires introduction of the size and the precision of the device for observing the field. Thus, the smallest scale is the size of the largest device of finite precision, which can be used to ''completely'' observe the field.

[24] A closure problem is actually encountered at multiple steps in the homogenization process. Thus, for $u(x)$ identified with the displacement field in the solid frame, the gradient of $u(x)$ is a measure of the local strain field in frame. A closure problem arises in relating the MAF defined on the local strain field to the gradient of the MAF defined on the local displacement field. Then, for $u(x)$ identified with the stress field in one of the phases, the gradient of $u(x)$ is a measure of the local linear momentum in that phase. A closure problem arises in writing a momentum balance equation expressed in MAFs defined on the stress and displacement fields.

[25] J. C. Slattery, ''Flow of viscoelastic fluids through porous media,'' Am. Inst. Chem. Eng. J. **13**, 1066–1071 (1967).

[26] A similarity might be noted between the partitioning presented and one familiar to structural mechanics modeling, whereby a structural system is resolved in finite elements with responses that are grouped as ''primary'' and as ''secondary'' measures, depending on an external forcing ''acting'' or ''not acting'' on the finite element.

Acoustic properties of rarefied gases inside pores of simple geometries

Vitaly F. Kozlov[a] and Alexander V. Fedorov
Department of Aeromechanics and Flight Engineering, Moscow Institute of Physics and Technology, Gagarin Str., 16, Zhukovsky, Moscow Reg., Russia, 140187

Norman D. Malmuth
Rockwell Scientific Company, Thousand Oaks, California 91360

(Received 21 August 2004; revised 1 February 2005; accepted 28 February 2005)

Analytical solutions describing propagation of monochromatic acoustic waves inside long pores of simple geometries and narrow flat slits are obtained with accounting for gas rarefaction effects. It is assumed that molecular nature of gas is important in Knudsen layers near solid boundaries. Outside the Knudsen layers, the continuum approach is used. This model allows for extension of acoustic analysis to regions of low pressures and microscopic cross-sectional sizes of channels. The problem is solved using linearized Navier-Stokes equations with the boundary conditions that resulted from the first-order approximation with respect to small Knudsen number Kn. For slits and pores of circular and square cross sections, the theoretical dependencies of the dynamic density in the low-frequency range are compared with those that resulted from known experimental data on steady-state flows of rarefied gases in uniform channels. Despite the formal restriction $\mathrm{Kn} \ll 1$ of asymptotic analysis, the theoretical model agrees well with experiments up to $\mathrm{Kn} \sim 5$. It is shown that the molecular phenomena affect acoustic characteristics of micro-channels and pores starting from relatively small Knudsen numbers $\mathrm{Kn} > 0.01$, especially at low frequencies. The obtained results may be used for analyses of acoustic properties of waveguides, perforated panels, micro-channels and pores in wide range of gas pressures as well as for stationary flows of rarefied gases through long uniform pipes etc.
[DOI: 10.1121/1.1893428]

PACS numbers: 43.20.Bi, 43.20.Jr, 43.20.Mv, 43.55.Ev [RR] Pages: 3402–3412

I. INTRODUCTION

The classic solutions[1,2] of Navier-Stokes equations, which describe acoustic waves traveling inside a long circular tube, are widely used for analyses of acoustic properties of waveguides and perforated panels. They also form a foundation for the semi-empirical models[3–9] predicting the absorption of sound by rigid frame porous media of complex structures. The exact solutions describing acoustic waves inside pores of rectangular and triangular cross-sectional shapes were presented in Refs. 10–13. These results have been obtained with the assumption that gas inside pores is a continuum.

However, there are practical cases in which the continuum assumption is not valid. In particular, rarefaction effects become important in flows within porous materials, which have ultra-fine structure and/or are exploited at so low ambient pressures that the molecular mean free path λ is comparable to a characteristic length scale. For example, in the ultrasonically absorptive coating[14,15] designed for the laminar flow control at hypersonic speeds, the mean free path was approximately 40% of the pore radius $r_p = 25\,\mu$m. This leads to 20% increase of the boundary-layer stabilization effect produced by the coating.[15] Similar situations occur in microfluids[16] dealing with gas flows in capillaries and ducts of $\sim$1-μm diameter. Many of these cases may be modeled using the concept of slip flows, in which the gas is treated as a continuum except thin Knudsen layers on solid walls. Within the Knudsen layer of thickness $\sim\lambda$, the thermodynamic equilibrium is violated. As shown in the kinetic theory of rarefied gases, slip flows are governed by the Navier-Stokes equations with the boundary conditions that resulted from the first-order (instead of zeroth-order) approximation with respect to the small Knudsen number $\mathrm{Kn} = \lambda / L_p$, where L_p is a macroscopic length scale.

Kundt and Warburg[17] were the first who revealed experimentally the slip effects in the gas outflow from a circular tube. They showed that the volume rate of laminar low-density flow inside a circular tube exceeded that predicted by the Poiseuille law. Maxwell[18] gave a qualitative explanation of this phenomenon by means of molecular-kinetic consideration. Smoluchowski[19] discovered physical reasons causing the temperature discontinuity on the solid wall submerged into nonuniformly heated gas of low density. These effects were thoroughly studied both experimentally and theoretically. Detailed reviews of the main results are given in Refs. 20–22 and many other monographs dealing with the interaction of molecular gas with solid surfaces.

Similar to the case of stationary flows through pipes, the gas rarefaction may affect acoustic disturbances propagating in pores. To our knowledge, this problem has not yet been studied. This motivated us to analyze propagation of sound in isolated long pores of various cross-sectional shapes and determine their acoustic characteristics accounting for the

[a]Electronic mail: afedorov@pt.comcor.ru

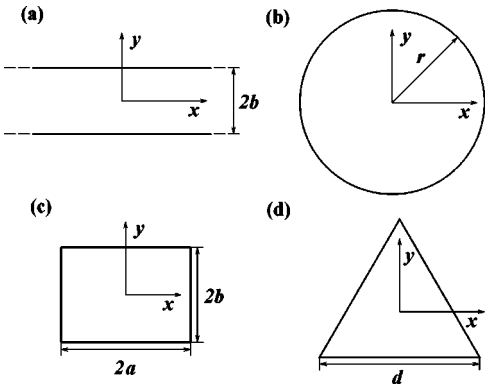


FIG. 1. The pore cross-sectional shapes. They are (a) slit of half-height b, (b) circle of radius r_0, (c) rectangle of half-width a and half-height b, and (d) equilateral triangle with sides d.

rarefaction effects. The acoustic problem for straight, uniform, and infinitely long pores is formulated in Sec. II. Its analytical solutions for flat slits and pores of circular, rectangular, and triangular cross-sections are obtained in Section III. The theoretical results are compared with available experimental data in Sec. IV and summarized in Sec. V.

II. FORMULATION OF THE PROBLEM

Consider a pulsating flow of rarefied gas inside flat slits and straight uniform pores. It is assumed that the rarefaction effects are appreciable in thin layers (Knudsen layers) forming on the solid boundaries. Outside the Knudsen layers, the fluid is treated as a continuum medium. The corpuscular effects are taken into account via boundary conditions. In the kinetic theory (see, for example, Refs. 21 and 22) this model corresponds to the first-order asymptotic approximation with respect to small Knudsen number Kn. As usual for asymptotic models, the range of Knudsen numbers in which the first-order approximation is robust can be estimated by comparison with experimental data.

The Cartesian coordinate system (x,y,z) is used for pores of arbitrary cross-sectional shape, and the cylindrical system (r,φ,z) is used for circular pores, where z is measured along the longitudinal axis of the pore. The origin of each coordinate system is located at the cross-sectional center of the pore as sketched in Fig. 1. With the assumption that the longitudinal pore length is much larger than its cross-sectional size L_p, the end effects are neglected. Gas motions inside the pore are caused by an infinitesimal periodic pressure gradient $\mathrm{Re}\{(d\tilde{p}/dz)\exp(i\omega t)\}$, where $i=\sqrt{-1}$, $\omega=2\pi f$ is angular frequency, t is time, and $d\tilde{p}/dz$ is complex amplitude. Solutions of the linearized Navier-Stokes equations for the particle velocity u and temperature disturbances T' are expressed in the form $u=\tilde{u}\exp(i\omega t)$ and $T'=\tilde{T}\exp(i\omega t)$. The momentum and energy equations lead to the equations for complex amplitudes $\tilde{u}$ and $\tilde{T}$:

$$\Delta\tilde{u}+\Lambda^2\tilde{u}=-\frac{\Lambda^2}{i\omega\rho_0}\frac{d\tilde{p}}{dz}, \tag{1a}$$

$$\Delta\tilde{T}+\mathrm{Pr}\Lambda^2\tilde{T}=\frac{\mathrm{Pr}\Lambda^2}{\rho_0 C_\mathrm{p}}\tilde{p}, \tag{1b}$$

where Pr is Prandtl number ($\mathrm{Pr}=\eta C_\mathrm{p}/\kappa$), C_p is the specific heat capacity under constant pressure, η and κ are the dynamic viscosity and heat conductivity of gas, respectively; $\Lambda=\sqrt{-i\omega\rho_0 L_p^2/\eta}$ is the dimensionless parameter, the modulus of which is the ratio of the characteristic length L_p to the viscous skin depth $\delta_\mathrm{visc}=\sqrt{\eta/\omega\rho_0}$, and ρ_0 is gas density in equilibrium; the Laplacian $\Delta=\partial^2/\partial y^2$ for slits, $\Delta=(1/r)(\partial/\partial r)(r\partial/\partial r)$ for the pore of circular cross section, and $\Delta=\partial^2/\partial x^2+\partial^2/\partial y^2$ for pores of rectangular and triangular cross-sections. The coordinates x, y, r are dimensionalized using L_p. The cross-sectional length scale L_p is specified as: $L_p=b$ is the half-height of the slit; $L_p=r_0$ is the radius of the cylindrical pore; $L_p=\sqrt{3}d/4$ is the half-height of the equilateral triangle. For the rectangular pore, the characteristic size, L_C, may be determined as the pore half-width a or half-height b, the diagonal $2\sqrt{a^2+b^2}$, the hydraulic radius $2ab/(a+b)$ depending on applications. Thus, the key macroscopic parameter Λ is specified as

$$\Lambda=\begin{cases}\sqrt{-i\omega\rho_0 b^2/\eta}, & \text{slit;}\\ \sqrt{-i\omega\rho_0 r_0^2/\eta}, & \text{circle;}\\ \sqrt{-i\omega\rho_0 L_C^2/\eta}, & \text{rectangle;}\\ \sqrt{-3i\omega\rho_0 d^2/16\eta}, & \text{triangle.}\end{cases} \tag{2}$$

In accordance with the concept of slip flow, the boundary conditions on the solid surfaces are written as[23] (in our notations)

$$u_g=B_u\left(\frac{\partial u}{\partial N}\right)_w, \tag{3a}$$

$$T_g=B_E\left(\frac{\partial T}{\partial N}\right)_w, \tag{3b}$$

where $\partial/\partial N$ is the directional derivative calculated along the unit normal drawn into the pore, the subscript "w" denotes quantities on the pore wall, and u_g and T_g are perturbations of gas velocity and temperature on the wall surface. The factors B_u and B_E are introduced for brevity. They are expressed in terms of the molecular tangential impulse α_u and energy α_E accommodation coefficients, the specific heat ratio γ, Prandtl number Pr and Knudsen number Kn, as

$$B_u=(2\alpha_u^{-1}-1)\mathrm{Kn},\quad B_E=(2\gamma/\mathrm{Pr}(\gamma+1))(2\alpha_E^{-1}-1)\mathrm{Kn}.$$

The accommodation coefficients α_u and α_E depend, mainly, on collision interaction laws between gas molecules and the solid wall, and on gas temperature, and depend almost not at all on pressure (see, for example, Ref. 20). In practice, an approximation $\alpha_u\approx\alpha_E$ is often used.

Hereafter, we use the well-known theoretical expression for the dynamic viscosity of monatomic gas modeled by hard spheres[23] $\eta\approx 0.5\rho_0\bar{c}\lambda$ to express the molecular mean free path λ through experimentally determined parameters ρ_0, T_0, and η. Accordingly, the Knudsen number is specified as $\mathrm{Kn}\approx 2\eta/(\rho_0\bar{c}L_p)$, where $\bar{c}=\sqrt{8R_gT_0/\pi}$ is the mean molecular speed, R_g is gas constant per unit mass, and T_0 is gas temperature in equilibrium. In addition, the unperturbed solid surface temperature T_w in (3b) is assumed to be constant.

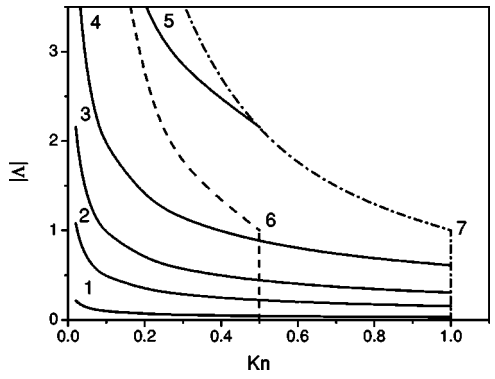


FIG. 2. The $|\Lambda|$-Kn diagram of the regimes of validity of the asymptotic theory. Isolines 1–5 correspond to constant ratios $L_p/\lambda\!\!\!^{-}$: $1-L_p/\lambda\!\!\!^{-}=10^{-4}$, $2-L_p/\lambda\!\!\!^{-}=2.5\cdot10^{-3}$, $3-L_p/\lambda\!\!\!^{-}=10^{-2}$, $4-L_p/\lambda\!\!\!^{-}=0.04$, and $5-L_p/\lambda\!\!\!^{-}=0.25$. The dashed curve 6 restricts the parameter plane by $\mathrm{Kn}<0.5$ in the low-frequency range and by $\mathrm{Kn}|\Lambda|<0.5$ in the high-frequency range, respectively. The dash-and-dot line 7 does the same for the corresponding restrictions $\mathrm{Kn}<1$ and $\mathrm{Kn}|\Lambda|<1$.

The boundary conditions (3a) and (3b), which are obtained in the first-order approximation with respect to the small ratio of molecular mean free path λ to a macroscopic length scale, give a nonzero relative macroscopic gas velocity and temperature jump on the solid wall. More general forms of these conditions, in particular, for varying wall temperature T_w, are discussed, for example, in Refs. 21 and 22.

In the problem considered herein, there are two important macroscopic length scales: the cross-sectional pore size L_p and the viscous skin depth δ_{visc}. The asymptotic analysis should be conducted using a small parameter to be determined as the largest of ratios λ/L_p and $\lambda/\delta_{\mathrm{visc}}$. In the low-frequency range ($|\Lambda|<1$) the viscous skin depth δ_{visc} exceeds the cross-sectional size L_p, and the small parameter is the Knudsen number $\mathrm{Kn}=\lambda/L_p$. In the high-frequency range ($|\Lambda|>1$) the characteristic cross-sectional size L_p is larger than δ_{visc}, and the proper small parameter is $\mathrm{Kn}|\Lambda|$ (assuming that $\mathrm{Pr}\sim1$). Figure 2 illustrates these two cases for air ($\gamma=1.4$) in the parameter space $\mathrm{Kn}-|\Lambda|$. Using definition of Kn and (2) we obtain the relation $|\Lambda|=(\pi\gamma/2)^{1/4}\sqrt{kL_p/\mathrm{Kn}}$, where $k=2\pi/\lambda\!\!\!^{-}$ is the wave number and $\lambda\!\!\!^{-}=f^{-1}\sqrt{\gamma R_g T_0}$ is the wavelength of a monochromatic acoustic wave propagating in unbounded gas medium. With the help of this relation, we plotted the isolines $L_p/\lambda\!\!\!^{-}=\mathrm{const}$ (solid lines). The dashed line 6 and dash-dot line 7 indicate restrictions on the asymptotic model. From the left of line 6 the parameters satisfy the restrictions: $\mathrm{Kn}<0.5$ in low-frequency range, $\mathrm{Kn}|\Lambda|<0.5$ in high-frequency range. The line 7 corresponds to the restrictions $\mathrm{Kn}<1$ and $\mathrm{Kn}|\Lambda|<1$. This diagram allows us to estimate the wavelength (and frequency) range in which the theory is applicable. Say the molecular mean free path is $\lambda=0.6\times10^{-5}$ cm, which is typical for air at normal ambient conditions in accordance with a model of hard spheres.[23] For the pore cross-sectional size $L_p=3\times10^{-5}$ cm, the Knudsen number is $\mathrm{Kn}=0.2$. Drawing the vertical line $\mathrm{Kn}=0.2$ up to the intersection, for example, with the boundary 6 one finds that the dimensionless parameter $|\Lambda|$ should not exceed 2.6. This corresponds to the isoline $L_p/\lambda\!\!\!^{-}\approx1/7$ on which $\lambda\!\!\!^{-}\approx2$ μm is treated as a minimal allowable wavelength. For air temperature $T_0\approx273$ K, the corresponding maximal frequency is $f\sim10^8$ Hz. Since this frequency is far beyond the ultrasonic range, the theory is applicable to any practical range of acoustic disturbances.

The boundary conditions (3) cannot be satisfied for a polygonal cross section, since the direction of the surface normal N jumps at the polygon apices. This may cause sudden changes of the tangential viscous stress and the normal heat flux on the pore surface. To avoid this unphysical behavior, additional restrictions must be imposed, namely, continuity of the shear stresses and heat fluxes acting on the fracture lines of the pore surface. Analysis of the solutions[10,11,13] for rectangular and triangular pores shows that theoretical values of the tangential viscous stresses and the normal heat flux vanish at the polygonal vertices. Obviously, solutions considered herein should also possess this property for $\mathrm{Kn}\to0$. Accordingly, the dominant approximation gives

$$\left(\frac{\partial F}{\partial N}\right)_{(x_w,y_w)\to(x_s-0,y_s-0)}=\left(\frac{\partial F}{\partial N}\right)_{(x_w,y_w)\to(x_s+0,y_s+0)},$$

$$\left(\frac{\partial F}{\partial N}\right)_{w(x_w,y_w)\to(x_s\pm0,y_s\pm0)}\longrightarrow O(\Lambda^2)\times\mathrm{Kn} \tag{4}$$

where subscript "s" denotes the sth salient point of the boundary and F stands for the function u or T'.

III. SOLUTION OF THE PROBLEM

At first, we consider the cases of slit, circular, and triangular pores. The solution for rectangular pores will be obtained later in a different manner. We perform the following substitutions in (1):

$$\tilde{u}(x,y,z)=-\frac{1}{i\omega\rho_0}\frac{\mathrm{d}\tilde{p}}{\mathrm{d}z}[1-F(x,y,B_u,\Lambda)], \tag{5}$$

$$\tilde{T}(x,y,z)=\frac{\tilde{p}}{\rho_0 C_{\mathrm{p}}}[1-F(x,y,B_E,\tilde{\Lambda})], \tag{6}$$

where $\tilde{\Lambda}=\sqrt{\mathrm{Pr}}\Lambda$. In accordance with (1) and (3), the new unknown dimensionless functions $F(x,y,B_u,\Lambda)$ and $F(x,y,B_E,\tilde{\Lambda})$ satisfy the uncoupled Helmholtz equations in two-dimensional space (except the case of slit),

$$(\Delta+\Lambda^2)F(x,y,B_u,\Lambda)=0, \tag{7a}$$

$$(\Delta+\tilde{\Lambda}^2)F(x,y,B_E,\tilde{\Lambda})=0, \tag{7b}$$

and the corresponding boundary conditions

$$F(x_w,y_w,B_u,\Lambda)-B_u\left(\frac{\partial F(x,y,B_u,\Lambda)}{\partial N}\right)_w=1, \tag{8a}$$

$$F(x_w,y_w,B_E,\tilde{\Lambda})-B_E\left(\frac{\partial F(x,y,B_E,\tilde{\Lambda})}{\partial N}\right)_w=1. \tag{8b}$$

If the boundary contains salient points (the case of pores of triangular cross sections), then the functions F must satisfy the additional conditions (4). Since the problems (7a)–(8a), and (7b)–(8b), are similar, it is sufficient to solve the

first one. The solution of the second problem is obtained by the formal substitutions $B_u \to B_E$ and $\Lambda \to \tilde{\Lambda}$.

In the continuum approximation, the problem for pores of rectangular and triangular cross sections was solved in Refs. 10–13 using Fourier series. However, these series are not differentiable in the vicinity of solid boundaries, because they include decompositions of discontinuous functions. Since the boundary condition (8a) contains the partial derivatives with respect to the space variables, this method cannot be applied to the problem (7a) and (8a). This motivated us to use the appropriate eigensolutions of (7a):

$$F(x,y,B_u,\Lambda)=\begin{cases} A\cos\Lambda y, & \text{slit;} \\ BJ_0(\Lambda r), & \text{circle;} \\ \sum_{i=1}^{2} a_i\varphi_i(x,y,\Lambda), & \text{triangle;} \end{cases} \tag{9}$$

where A, B, a_1, and a_2 are constants of integration; $J_0(\Lambda r)$ is Bessel function of the first kind of order zero; $\varphi_1(x,y,\Lambda)=\sin\Lambda\xi_1+\sin\Lambda\xi_2-\sin\Lambda\xi_3$ and $\varphi_2(x,y,\Lambda)=\cos\Lambda\xi_1+\cos\Lambda\xi_2+\cos\Lambda\xi_3$ are independent eigensolutions of (7a), $\xi_1=y/2+\sqrt{3}x/2$, $\xi_2=y/2-\sqrt{3}x/2$, $\xi_3=y$.

In the cases of slit and circular pores, the coefficients A and B are solutions of the algebraic equation resulted from substitution of (9) into the boundary condition (8a):

$$A=\frac{1}{\cos\Lambda\{1-B_u\Lambda\tan\Lambda\}}, \tag{10}$$

$$B=\frac{1}{J_0(\Lambda)\{1-0.5B_u\Lambda^2Q(\Lambda)\}}, \tag{11}$$

where $Q(\Lambda)=2J_1(\Lambda)/\Lambda J_0(\Lambda)$.

Owing to the rotation invariance of the solution for triangular pore, it is sufficient to consider the boundary conditions on one side of the triangle, say $y=-\frac{2}{3}$. Substituting the corresponding expressions into (8a) at $y=-\frac{2}{3}$ and equating to zero the coefficients of similar terms, we obtain the linear algebraic system for a_1 and a_2:

$$a_1\sin\frac{2\Lambda}{3}\left(1+\Lambda B_u\cot\frac{2\Lambda}{3}\right)+a_2\cos\frac{2\Lambda}{3}\left(1-\Lambda B_u\tan\frac{2\Lambda}{3}\right)=1,$$

$$a_1\sin\frac{\Lambda}{3}\left(1+\frac{\Lambda B_u}{2}\cot\frac{\Lambda}{3}\right)-a_2\cos\frac{\Lambda}{3}\left(1-\frac{\Lambda B_u}{2}\tan\frac{\Lambda}{3}\right)=0.$$

Its solution is expressed as

$$a_1=\cos\frac{\Lambda}{3}\left(1-\frac{\Lambda}{2}B_u\tan\frac{\Lambda}{3}\right)\Big/\Delta_u,$$

$$a_2=\sin\frac{\Lambda}{3}\left(1+\frac{\Lambda}{2}B_u\cot\frac{\Lambda}{3}\right)\Big/\Delta_u, \tag{12}$$

where $\Delta_u=[1+1.5\Lambda B_u\cot\Lambda-0.5(\Lambda B_u)^2]\sin\Lambda$. Direct substitution shows that the solution (9) and (12) satisfies the condition (8a) and the additional conditions (4) at the triangle apices.

Now we consider the case of the rectangular pore. Note that the two-dimensional Fourier expansion

$$\tilde{u}(x,y,z)=\frac{4\Lambda^2}{i\omega\rho_0}\frac{d\tilde{p}}{dz}\times\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(-1)^{m+n}\cos\alpha_m q_1x\cos\alpha_n q_2y}{\alpha_m\alpha_n(\alpha_m^2q_1^2+\alpha_n^2q_2^2-\Lambda^2)},$$

which was used in Refs. 10 and 11 for continuous media, is reduced to the one-dimensional form using the Fourier series

$$1-\frac{\cosh(\beta_{2m}q_1x)}{\cosh\beta_{2m}}=2\beta_{2m}^2\sum_{n=0}^{\infty}\frac{(-1)^n\cos\alpha_nq_1x}{\alpha_n(\alpha_n^2+\beta_{2m}^2)},$$

$$1-\frac{\cosh(\beta_{1m}q_2y)}{\cosh\beta_{1m}}=2\beta_{1m}^2\sum_{n=0}^{\infty}\frac{(-1)^n\cos\alpha_nq_2y}{\alpha_n(\alpha_n^2+\beta_{1m}^2)},$$

where $(|q_1x|\leqslant 1,|q_2y|\leqslant 1)$. This form is written as

$$\tilde{u}(x,y,z)=\frac{\Lambda^2}{i\omega\rho_0}\frac{d\tilde{p}}{dz}\sum_{m=0}^{\infty}\frac{(-1)^m}{\alpha_m}\left\{\frac{1}{q_2^2\beta_{1m}^2}\left(1-\frac{\cosh(\beta_{1m}q_2y)}{\cosh\beta_{1m}}\right)\times\cos(\alpha_mq_1x)+\frac{1}{q_1^2\beta_{2m}^2}\left(1-\frac{\cosh(\beta_{2m}q_1x)}{\cosh\beta_{2m}}\right)\times\cos(\alpha_mq_2y)\right\}, \tag{13}$$

where $q_1^{-1}=a/L_C$ and $q_2^{-1}=b/L_C$ are dimensionless rectangle sides, $\beta_{1m}=q_2^{-1}\sqrt{\alpha_m^2q_1^2-\Lambda^2}$, $\beta_{2m}=q_1^{-1}\sqrt{\alpha_m^2q_2^2-\Lambda^2}$, and $\alpha_m=(m+0.5)\pi$. Note that we use the two independent complete systems of functions to symmetrize (13) with respect to x and y. Now we can modify (13) in order to satisfy the boundary condition (3a). This modification is written as

$$\tilde{u}(x,y,z)=\frac{\Lambda^2}{i\omega\rho_0}\frac{d\tilde{p}}{dz}\sum_{m=0}^{\infty}\sum_{j=1}^{2}\{a_{jm}\Psi_{jm}(x,y,B_u,\Lambda)\},$$

$$\Psi_{1m}(x,y,B_u,\Lambda)=\frac{(-1)^m}{q_2^2\gamma_{1m}\beta_{1m}^2}\left(1-\frac{\cosh(\beta_{1m}q_2y)}{Q_{1m}\cosh\beta_{1m}}\right)\cos(\gamma_{1m}q_1x),$$

$$Q_{1m}=1+q_2B_u\beta_{1m}\tanh\beta_{1m},$$

$$\Psi_{2m}(x,y,B_u,\Lambda)=\frac{(-1)^m}{q_1^2\gamma_{2m}\beta_{2m}^2}\left(1-\frac{\cosh(\beta_{2m}q_1x)}{Q_{2m}\cosh\beta_{2m}}\right)\cos(\gamma_{2m}q_2y), \tag{14}$$

$$Q_{2m}=1+q_1B_u\beta_{2m}\tanh\beta_{2m},$$

$$\beta_{1m}=q_2^{-1}\sqrt{\gamma_{1m}^2q_1^2-\Lambda^2},\quad \beta_{2m}=q_1^{-1}\sqrt{\gamma_{2m}^2q_2^2-\Lambda^2},$$

where a_{jm} $(j=1,2)$ are unknown integration constants; the eigenvalues γ_{1m} and γ_{2m} are roots of the equations

$$\cos\gamma_{1m}=q_1B_u\gamma_{1m}\sin\gamma_{1m}, \tag{15a}$$

$$\cos\gamma_{2m}=q_2B_u\gamma_{2m}\sin\gamma_{2m}. \tag{15b}$$

In the expansion (14), each function $\Psi_{jm}(x,y,B_u,\Lambda)$ $(j=1,2)$ satisfies the boundary condition (3a). Hereafter we

explain how to find the coefficients a_{jm} in the decomposition (14). Direct substitution shows that (14) satisfies (1a), if $\Sigma_{m=0}^{\infty}(-1)^m\{a_{1m}\gamma_{1m}^{-1}\cos(\gamma_{1m}q_1x)+a_{2m}\gamma_{2m}^{-1}\cos(\gamma_{2m}q_2y)\}=1$. The left-hand side of this equality may be interpreted as a sum of two series: the first series depends on x and the second on y. The sum of these two terms may be constant if and only if each of them is constant. Thus, taking into account the solution (13) for continuous medium we write

$$\sum_{m=0}^{\infty}(-1)^m a_{1m}\gamma_{1m}^{-1}\cos(\gamma_{1m}q_1x)=\frac{1}{2},$$
$$\sum_{m=0}^{\infty}(-1)^m a_{2m}\gamma_{2m}^{-1}\cos(\gamma_{2m}q_2y)=\frac{1}{2}. \tag{16}$$

This system of linear algebraic equations has a simple solution at $B_u=0$. In this case the system (15) gives $\gamma_{1m}=\gamma_{2m}=\alpha_m=(m+0.5)\pi$. Using Fourier decompositions we obtain the equation $2\Sigma_{m=0}^{\infty}(-1)^m\alpha_m^{-1}\cos(\alpha_m q_1x)=2\Sigma_{m=0}^{\infty}(-1)^m\alpha_m^{-1}\cos(\alpha_m q_2y)=1$ ($|q_1x|<1, |q_2y|<1$), which gives $a_{1m}=a_{2m}=1$ ($m=0,1,2,...$). Thus, the solution (14) coincides with (13) in the case of $B_u=0$. Direct verification shows that (14) satisfies the conditions (4) at the angular points of rectangle.

The functions Ψ_{1m} and Ψ_{2m}, which are used in the case of rectangular pore, do not represent an orthogonal system on the rectangle sides. For this reason we cannot immediately solve the system (16). The orthogonalization is performed using the complete trigonometric systems $\{\cos\alpha_n q_1x\}$ and $\{\cos\alpha_n q_2y\}$. Substituting the decompositions

$$\cos(\gamma_{1m}q_1x)-\cos\gamma_{1m}=2\gamma_{1m}^2\cos\gamma_{1m}\sum_{n=0}^{\infty}\frac{(-1)^n\cos\alpha_n q_1x}{\alpha_n(\alpha_n^2-\gamma_{1m}^2)}\quad(|q_1x|\leqslant 1),$$

$$\cos(\gamma_{2m}q_2y)-\cos\gamma_{2m}=2\gamma_{2m}^2\cos\gamma_{2m}\sum_{n=0}^{\infty}\frac{(-1)^n\cos\alpha_n q_2y}{\alpha_n(\alpha_n^2-\gamma_{2m}^2)}\quad(|q_2y|\leqslant 1)$$

into (16) and using the orthogonal properties of the trigonometric functions we reduce the problem to solving two uncoupled infinite systems of algebraic equations ($j=1,2$):

$$\sum_{n=0}^{\infty}\frac{\cos(\gamma_{jn}-n\pi)}{\gamma_{jn}(\alpha_m^2-\gamma_{jn}^2)}a_{jn}=\frac{1}{2\alpha_m^2}. \tag{17}$$

From (15) we also obtain

$$\cos(\gamma_{jn}-n\pi)=\frac{q_jB_u\gamma_{jn}}{\sqrt{1+(q_jB_u\gamma_{jn})^2}}\quad(j=1,2).$$

Owing to this we can express (17) in the form

$$\sum_{n=0}^{\infty}\frac{a_{jn}}{(\alpha_m^2-\gamma_{jn}^2)\sqrt{1+(q_jB_u\gamma_{jn})^2}}=\frac{1}{2q_jB_u\alpha_m^2}. \tag{18}$$

As discussed above, the solution of (1b) is obtained by the replacement $(B_u,\Lambda)\to(B_E,\tilde{\Lambda})$ in the solution of (1a). This leads to the amplitude of temperature pulsations

$$\tilde{T}(x,y,z)=-\frac{\tilde{\Lambda}^2\tilde{p}}{\rho_0C_p}\sum_{m=0}^{\infty}\sum_{j=3}^{4}\{a_{jm}\Psi_{jm}(x,y,B_E,\tilde{\Lambda})\}, \tag{19}$$

where

$$\Psi_{3m}(x,y,B_E,\tilde{\Lambda})=\frac{(-1)^m}{q_2^2\gamma_{3m}\beta_{3m}^2}\left(1-\frac{\cosh(\beta_{3m}q_2y)}{Q_{3m}\cosh\beta_{3m}}\right)\cos(\gamma_{3m}q_1x),$$
$$Q_{3m}=1+q_2B_E\beta_{3m}\tanh\beta_{3m},$$

$$\Psi_{4m}(x,y,B_E,\tilde{\Lambda})=\frac{(-1)^m}{q_1^2\gamma_{4m}\beta_{4m}^2}\left(1-\frac{\cosh(\beta_{4m}q_1x)}{Q_{4m}\cosh\beta_{4m}}\right)\cos(\gamma_{4m}q_2y),$$
$$Q_{4m}=1+q_1B_E\beta_{4m}\tanh\beta_{4m},$$

$$\beta_{3m}=q_2^{-1}\sqrt{\gamma_{3m}^2q_1^2-\tilde{\Lambda}^2},\quad \beta_{4m}=q_1^{-1}\sqrt{\gamma_{4m}^2q_2^2-\tilde{\Lambda}^2}.$$

Equations for the eigenvalues γ_{jm} and coefficients a_{jm} of the solutions (14) and (19) may be written in the common form for $j=1, 2, 3, 4$; $m=0,1,2,...$:

$$\cos\gamma_{jm}=h_j\gamma_{jm}\sin\gamma_{jm}, \tag{20}$$

$$\sum_{n=0}^{\infty}\frac{\cos(\gamma_{jn}-n\pi)}{\gamma_{jn}(\alpha_m^2-\gamma_{jn}^2)}a_{jn}=\frac{1}{2\alpha_m^2}, \tag{21}$$

where $h_1=q_1B_u$, $h_2=q_2B_u$, $h_3=q_1B_E$, and $h_4=q_2B_E$.

Approximate solutions of the equations (20) may be obtained as follows. Noting that $m\pi\leqslant\gamma_{jm}\leqslant(m+0.5)\pi$ we express the eigenvalues in the form

$$\gamma_{jm}=(m+0.25)\pi+\Delta\gamma_{jm}=\delta_m+\Delta\gamma_{jm},$$

which allows us to rearrange (20) as

$$\Delta\gamma_{jm}=\frac{\pi}{4}-\tan^{-1}(Z_{jm})+\frac{1}{2i}\left\{\ln\left(1-\frac{ih_j\Delta\gamma_{jm}}{1-iZ_{jm}}\right)-\ln\left(1+\frac{ih_j\Delta\gamma_{jm}}{1+iZ_{jm}}\right)\right\}$$
$$\approx\frac{\pi}{4}-\tan^{-1}(Z_{jm})-\frac{h_j\Delta\gamma_{jm}}{1+Z_{jm}^2}+\frac{Z_{jm}}{(1+Z_{jm}^2)^2}(h_j\Delta\gamma_{jm})^2+\frac{1-3Z_{jm}^2}{3(1+Z_{jm}^2)^3}(h_j\Delta\gamma_{jm})^3-\frac{Z_{jm}(1-Z_{jm}^2)}{(1+Z_{jm}^2)^4}(h_j\Delta\gamma_{jm})^4-\cdots,\quad Z_{jm}=h_j\delta_m.$$

This equation may be solved in various ways. An approximate solution may be obtained, for example, using the method of successive iterations

$$\Delta\gamma_{jm} \approx \frac{(1+Z_{jm}^2)\tau_{jm}}{1+h_j+Z_{jm}^2}\left\{1+\frac{h_j^2 Z_{jm}\tau_{jm}}{(1+h_j+Z_{jm}^2)^2}+\cdots\right\}, \quad (22)$$

where $\tau_{jm}=\pi/4-\tan^{-1}(Z_{jm})$. Numerical tests showed that the maximal relative deviation of the results obtained with use of (22) from corresponding numerical solutions was less then 10^{-3} at $m=0$ and Kn=1.0 ($q_j \leqslant 1$).

Since (20) and (21) do not depend on the parameter Λ (or $\tilde{\Lambda}$), the eigenvalues γ_{jm} and coefficients a_{jm} are determined once (for example, numerically) for particular q_1, q_2, B_u, and B_E. Then these solutions may be used for calculations of the acoustic characteristics at any dimensionless frequency Λ^2.

Since the pressure disturbance in the considered approach does not depend on the transversal variables (x,y), acoustic properties of an isolated pore can be obtained by averaging of the particle velocity and temperature excess over the pore cross-sectional area. In this connection, acoustics of porous media deals with the following parameters:[3,10]

(a) Dynamic density

$$\tilde{\rho}(B_u,\Lambda)=-\frac{1}{i\omega\langle\tilde{u}(z)\rangle}\frac{d\tilde{p}}{dz}. \quad (23)$$

(b) Dynamic compressibility

$$\tilde{C}(B_E,\gamma,\tilde{\Lambda})=\rho_0^{-1}(\langle\tilde{\rho}\rangle/\tilde{p}) = \frac{1}{p_0}-\frac{\langle\tilde{T}\rangle}{T_0\tilde{p}}=\frac{1}{p_0}\left(1-\frac{\rho_0 R\langle\tilde{T}\rangle}{\tilde{p}}\right), \quad (24)$$

where p_0 is unperturbed ambient pressure and $\langle\ \rangle$ denotes averaging over the cross-sectional area S, $\langle\tilde{f}\rangle = (1/S)\oint_S \tilde{f}(x,y)ds$. In turn, the dynamic density and dynamic compressibility are used to calculate the characteristic impedance Z_w and the propagation constant m:

$$Z_w=\sqrt{\tilde{\rho}/\tilde{C}}, \quad (25)$$

$$m=i\omega\sqrt{\tilde{\rho}\tilde{C}}. \quad (26)$$

Using (5) and (6) we obtain

$$\langle\tilde{u}(z)\rangle=-\frac{1}{i\omega\rho_0}\frac{d\tilde{p}}{dz}[1-\langle F(B_u,\Lambda)\rangle], \quad (27)$$

$$\langle\tilde{T}(z)\rangle=\frac{\tilde{p}}{\rho_0 C_p}[1-\langle F(B_E,\tilde{\Lambda})\rangle]. \quad (28)$$

Substituting (27) and (28) into (23) and (24) we express the complex dynamic density $\tilde{\rho}(B_u,\Lambda)$ and the dynamic compressibility $\tilde{C}(B_E,\gamma,\tilde{\Lambda})$ as

$$\tilde{\rho}(B_u,\Lambda)=\rho_0/[1-\langle F(B_u,\Lambda)\rangle], \quad (29)$$

$$\tilde{C}(B_E,\gamma,\tilde{\Lambda})=\frac{1}{\gamma p_0}[1+(\gamma-1)\langle F(B_E,\tilde{\Lambda})\rangle], \quad (30)$$

where

$$\langle F(B_u,\Lambda)\rangle = \begin{cases} \dfrac{\tan\Lambda}{\Lambda[1-B_u\Lambda\tan\Lambda]}, & \text{slit;} \\ \dfrac{Q(\Lambda)}{1-0.5B_u\Lambda^2 Q(\Lambda)}, & \text{circle;} \\ 1+\Lambda^2\displaystyle\sum_{m=0}^{\infty}(a_{1m}\langle\Psi_{1m}\rangle+a_{2m}\langle\Psi_{2m}\rangle), & \text{rectangle;} \\ \dfrac{3(1-\Lambda\cot\Lambda+0.5\Lambda^2 B_u)}{\Lambda^2[1+1.5\Lambda B_u\cot\Lambda-0.5(\Lambda B_u)^2]}, & \text{triangle;} \end{cases} \quad (31)$$

$$\langle F(B_E,\gamma,\tilde{\Lambda})\rangle = \begin{cases} \dfrac{\tan\tilde{\Lambda}}{\tilde{\Lambda}(1-B_E\tilde{\Lambda}\tan\tilde{\Lambda})}, & \text{slit;} \\ \dfrac{Q(\tilde{\Lambda})}{1-0.5B_E\tilde{\Lambda}^2 Q(\tilde{\Lambda})}, & \text{circle;} \\ 1+\tilde{\Lambda}^2\displaystyle\sum_{m=0}^{\infty}(a_{3m}\langle\Psi_{3m}\rangle+a_{4m}\langle\Psi_{4m}\rangle), & \text{rectangle;} \\ \dfrac{3(1-\tilde{\Lambda}\cot\tilde{\Lambda}+0.5\tilde{\Lambda}^2 B_E)}{\tilde{\Lambda}^2[1+1.5\tilde{\Lambda}B_E\cot\tilde{\Lambda}-0.5(\tilde{\Lambda}B_E)^2]}, & \text{triangle;} \end{cases} \quad (32)$$

$$\langle\Psi_{jm}\rangle=\frac{\sin(\gamma_{jm}-m\pi)}{D_j\gamma_{jm}^2\beta_{jm}^2}\left(1-\frac{\tanh\beta_{jm}}{\beta_{jm}Q_{jm}}\right);$$

$$D_j=q_2^2 \quad \text{at } j=1,3 \quad \text{and } D_j=q_1^2 \quad \text{at } j=2,4.$$

In the case of triangular pore, the following averages are used:

$$\langle\varphi_1(\Lambda)\rangle=\frac{3}{\Lambda}\left[\frac{1}{2\Lambda}\left(\sin\frac{4\Lambda}{3}+\sin\frac{2\Lambda}{3}\right)-\cos\frac{2\Lambda}{3}\right],$$

$$\langle\varphi_2(\Lambda)\rangle=\frac{3}{\Lambda}\left[\frac{1}{2\Lambda}\left(\cos\frac{2\Lambda}{3}-\cos\frac{4\Lambda}{3}\right)+\sin\frac{2\Lambda}{3}\right].$$

For the Knudsen number Kn=0, the formulas (29) and (30) together with (31) and (32) coincide with the corresponding expressions (which are presented, for example, in Ref. 13) for continuum media.

Results of numerical calculations with use of (29)–(32) are presented in Fig. 3 for a circular pore filled with air at Knudsen numbers 0, 0.05, 0.1, and 0.5. In these calculations, we used the approximations $\alpha_u=\alpha_E=0.9$. This assumption is consistent with the data presented in Table II and with data from other literature sources. It is seen that influence of rarefaction effects on acoustical characteristics of air becomes noticeable at Kn~0.01 and increases with the Knudsen number.

IV. COMPARISON WITH EXPERIMENT

Unfortunately, we could not find experimental data on sound propagating in uniform channels filled with low-density gases. However, it is known (see, for example, Ref.

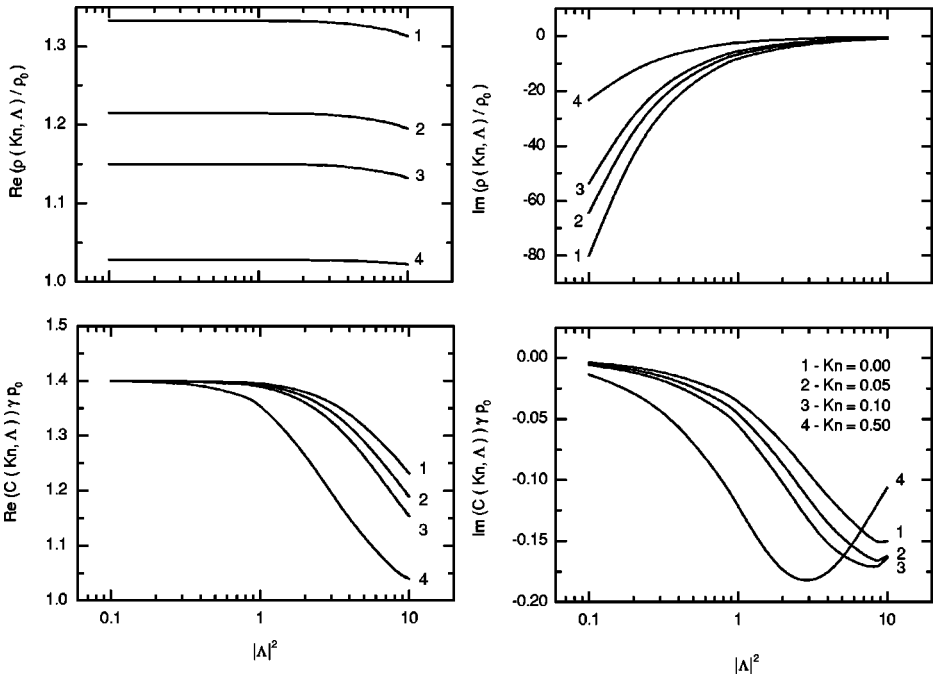


FIG. 3. Theoretical dependencies of the dynamic density and dynamic compressibility of air inside an infinitely long cylindrical pore via dimensionless parameters Λ and Kn, $\alpha_u = \alpha_E = 0.9$.

8) that in the low-frequency limit, $|\Lambda| \to 0$, acoustic properties of gas inside pores may be approximately expressed in terms of parameters characterizing steady viscous flows. For the considered above cases, this approximation may be also obtained using expansions of (31) and (32) into power series in Λ. For example, $\langle F(B_u, \Lambda) \rangle \approx 1 + (k_1 + cB_u)\Lambda^2 + \cdots$, where the decomposition coefficients k_1 and c depend on the pore cross-sectional shape: $k_1 = \frac{1}{3}$, $c = 1$ for slit; $k_1 = \frac{1}{8}$ and $c = 1/2$ for circular pores; etc. Substituting this decomposition into the exact formula (29) and using the definition of Λ one finds that $\tilde{\rho}(B_u, \Lambda) \underset{\Lambda \to 0}{\to} \sigma_0 / i\omega$, where $\sigma_0 = \eta/(k_1 + cB_u)L_p^2$ is resistivity of slip flow, which coincides with dc flow resistivity in the continuum approach $B_u \to 0$ (see also Ref. 13). Thus, the dynamic density is pure imaginary in the low-frequency limit. It is inversely proportional to frequency with the constant depending on stationary flow characteristics only. Note also that the rarefaction effect on acoustic properties is stronger in the low-frequency band owing to the relatively large imaginary part of the dynamic density, which tends to infinity as $\Lambda^2 \to 0$. This is also confirmed by dependencies shown in Fig. 3. All these arguments allow us to verify the theoretical model by comparisons with numerous experimental data on stationary flows of rarefied gases through long tubes of different cross-sections.

The definition (23) may be written in the form

$$\frac{\tilde{\rho}(B_u, \Lambda)}{\rho_0} = -\frac{1}{i\omega\rho_0 \langle \tilde{u}(z) \rangle} \frac{d\tilde{p}}{dz} = -\frac{S}{i\omega G} \frac{d\tilde{p}}{dz}, \tag{33}$$

where $G = \rho_0 \langle \tilde{u}(z) \rangle S$ is treated now as the stationary mass flow rate through a pore in the low-frequency limit. Normalizing the mass-flow rate as $\bar{G} = -G\sqrt{2R_g T_0}/SL_p(dp/dz)$, we express the dynamic density in terms of the dimensionless parameters

$$\bar{\rho} = \frac{\tilde{\rho}(B_u, \Lambda)}{\rho_0} = -\frac{\rho_0 L_p \sqrt{2R_g T_0}}{\eta \Lambda^2 \bar{G}} = -\frac{\sqrt{\pi}}{\Lambda^2 \mathrm{Kn} \bar{G}}. \tag{34}$$

For low frequencies ($|\Lambda| \to 0$), the dynamic compressibility is approximated as

$$\bar{C}(B_E, \gamma, \tilde{\Lambda}) = \gamma p_0 \tilde{C}(B_E, \gamma, \tilde{\Lambda}) \approx \gamma + O(B_E, \tilde{\Lambda}^2). \tag{35}$$

Formulas (33)–(35) allow us to evaluate $\tilde{\rho}$ and $\tilde{C}$ using parameters of the stationary flow through pores, and calculate the characteristic impedance Z_w and the propagation constant m using (25) and (26). The experimental data to be used for comparison were reported in forms, which differed from (33) and could not be directly substituted into (34). Hereafter we will derive a suitable modification of (34) for each source of data. All comparisons are performed for the imaginary part of dynamic density at $|\Lambda| = 10^{-1}$. The accommodation coefficient is $\alpha_u = 0.9$, besides the case of a slit for which α_u was measured in experiment. In this case, the relative error of the low-frequency approximation does not exceed 10^{-2}.

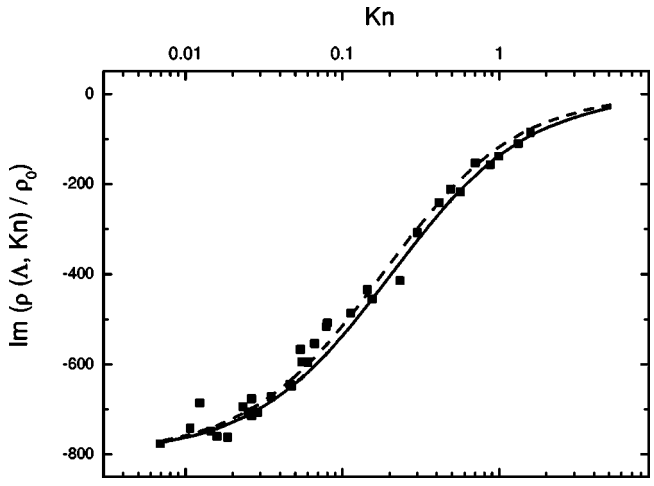


FIG. 4. Comparison of theoretical predictions (solid lines) for the normalized imaginary part of dynamic density at $|\Lambda| = 10^{-1}$, $\alpha_u = 0.9$ with those resulted from Knudsen's interpolation (36) (dashed lines) and experimental data[26] for flows of air in copper pipes of 1.30-cm radius and 332-cm length (squares).

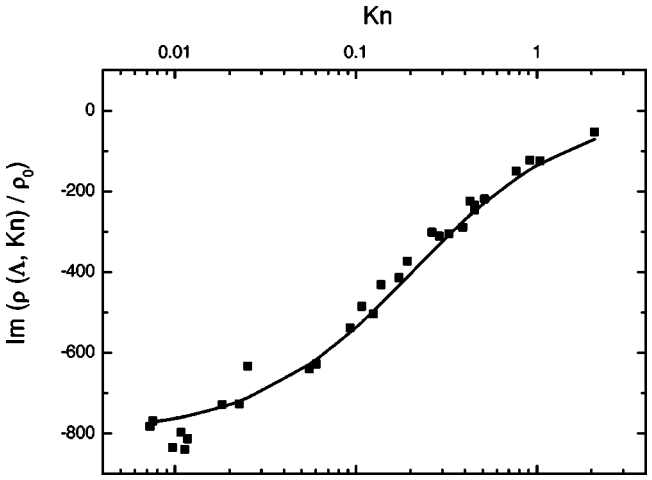


FIG. 5. Comparison of theoretical predictions (solid lines) for the normalized imaginary part of dynamic density at $|\Lambda|=10^{-1}$, $\alpha_u=0.9$ with those resulted from experimental data[26] for flows of hydrogen in copper pipes of 1.30-cm radius and 332-cm length (squares).

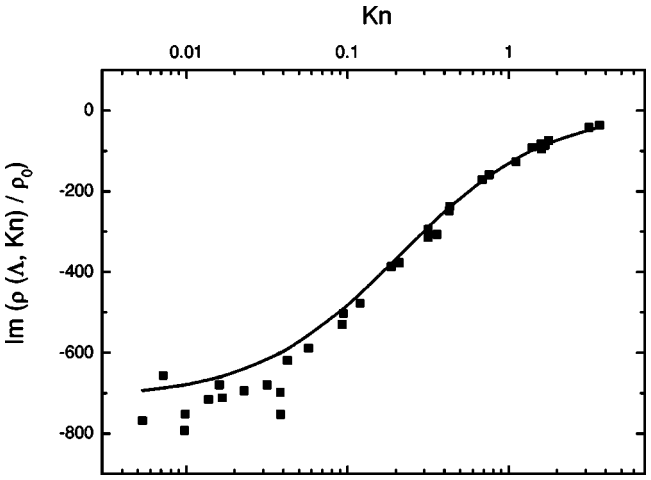


FIG. 6. Comparison of theoretical predictions (solid lines) for the normalized imaginary part of dynamic density at $|\Lambda|=10^{-1}$, $\alpha_u=0.9$ with those resulted from experimental data[27] for flow of argon in the square tube of 2.5-mm cross-section side and 305-mm length (squares).

A. Circular pores

The mass-flow rates of stationary flows in circular pipes were evaluated by many researchers, both theoretically and experimentally, for a wide range of Knudsen numbers (see, for example, Ref. 24). Some of the experimental data were approximated analytically. For example, the correlation (in our notation)

$$\overline{G}=\frac{\sqrt{\pi}}{8\mathrm{Kn}}\left(1+8.52\sqrt{\frac{2}{\pi}}\cdot\mathrm{Kn}\frac{\mathrm{Kn}+\sqrt{2\pi}}{\mathrm{Kn}+1.235\sqrt{2\pi}}\right) \tag{36}$$

is known as Knudsen's formula.[25] Along with (36) we use the data.[26] For comparison with these data, we write (34) in the form

$$\overline{\rho}=-i\frac{\pi}{\Lambda^2 XU}\left(\frac{r_0}{30.48}\right)^3, \tag{37}$$

where L is the pipe length in cm, r_0 is the pipe radius in cm, U is the gas flow rate per unit pressure drop in micron-cubic foot per second per micron, and $X=(\eta/r_0)p_m(\rho_1 g_c)^{1/2}$ is the dimensionless group of parameters defined in Ref. 26. The theoretical dependences of $\mathrm{Im}\,\tilde{\rho}(B_u,\Lambda)$ were calculated at $|\Lambda|=10^{-1}$ and compared with the experimental data in Figs. 4 and 5. The theoretical predictions (solid line) agree well with the Knudsen's interpolation (36) (dashed line) and with the data on air flow (squares in Fig. 4) and hydrogen flow (squares in Fig. 5) obtained in copper pipes of 1.30-cm radius and 332-cm length. The average relative deviation of the experimental data from the theory does not exceed 6% in the considered range of Knudsen numbers.

B. Square pipes

Data for pipes of square cross section were presented in Ref. 27. For straightforward usage of these data we express the dimensionless dynamic density in the form

$$\overline{\rho}=-\frac{101.325h^3\sqrt{2\pi}}{0.76\Lambda^2 N_K L m_{\mathrm{Ar}}\sqrt{R_{\mathrm{Ar}}T}(\dot{n}/\Delta p)}, \tag{38}$$

where $h\approx 1.25$ mm is the half-side of square, $L\approx 305$ mm is the tube length, m_{Ar} is the molecular mass of argon, and R_{Ar} is the gas constant per unit mass for argon. The Knudsen number N_K and the molecular flow rate per unit pressure $\dot{n}/\Delta p$ are tabulated in Ref. 27. The factor 101.325/0.76 is used to convert mm Hg to Pa.

The corresponding theoretical dependencies were calculated using (29)–(32) and (20) and (21) at $q_1=q_2=1$. The roots of (20) were determined using the method of bisection. Equations (21) were solved by the Gaussian method. For

TABLE I. Dimensionless flow rate $\overline{G}$ as a function of rarefaction parameter $\delta=2/\mathrm{Kn}$ (slit 3 in Ref. 28).

H_2		D_2		He		Ne		CO_2	
δ	$\overline{G}$	δ	$\overline{G}$	δ	$\overline{G}$	δ	$\overline{G}$	δ	$\overline{G}$
20.18	7.50	39.45	10.13	27.35	9.66	34.91	11.50	37.15	11.66
9.42	4.79	14.42	6.40	13.55	6.10	17.30	7.08	18.41	6.87
4.95	3.55	7.60	4.21	6.73	4.29	7.83	4.58	7.83	4.37
2.53	3.08	3.55	3.32	3.24	3.39	3.77	3.63	3.44	3.21
1.22	2.89	1.93	3.05	1.60	3.10	2.11	3.26	1.38	2.82
0.624	2.95	1.38	2.95	0.984	3.05	1.11	3.16	0.726	2.84
0.301	3.08	0.489	3.10	0.473	3.24	0.552	3.24	0.318	2.95
0.113	3.45	0.265	3.18	0.301	3.42	0.249	3.39	0.149	3.16
0.0578	3.71	0.113	3.58	0.144	3.60	0.116	3.74	0.0673	3.37

TABLE II. Slip constants $\sigma \pm \Delta\sigma$ (slit 3 in Ref. 28 or slit 2 in Ref. 29).

H_2	D_2	He	Ne	CO_2
1.374±0.014	1.407±0.014	1.494±0.010	1.558±0.017	1.235±0.017

each number j of the unknown coefficients, a system of 300 equations was solved. This provided a relative accuracy of the order of 10^{-8}.

Comparisons of the theoretical predictions (solid line) with the experimental data (squares) are shown in Fig. 6. The average relative deviation of experimental values from theoretical results is approximately 10% in the considered range of Knudsen numbers.

C. Narrow slits

The first experimental observations of rarefied gas flows through slits were made by Gaede.[28] However, for comparisons we use the later data,[29,30] which were obtained with accounting for roughness of slit walls and for which the mean accommodation coefficients were determined. Unfortunately, the experimental conditions of the results[29,30] were not well documented. In particular, the pressure gradient $d\tilde{p}/dz$ and the mean pressures inside slits were not specified. This leads to some uncertainty in evaluation of basic parameters such as Knudsen number. The dimensionless dynamic density is approximated using (34). The dimensionless flow rate G for different gases is given in Table I as a function of the rarefaction parameter $\delta = 2/\mathrm{Kn}$. These data were obtained by scanning of the plots in Fig. 5 of Ref. 29 for slit 3 (see also the data for slit 2 in Ref. 30).

Experimental values of the slip constant σ for various gases are given in Table II. They correspond to the data in Table II of Ref. 29. Since the slip constant is expressed via the molecular tangential impulse accommodation coefficient α_u as $\sigma = (2-\alpha_u)/\alpha_u$, then $B_u = \sigma\,\mathrm{Kn}$. Data from Table II were used for computations of the theoretical and experimental distributions $\mathrm{Im}\,\tilde{\rho}(B_u, \Lambda)$.

Results of comparison are shown in Figs. 7–9. Since the dimensionless theoretical distributions weakly depend on Prandtl number and do not depend on other individual gas parameters (besides the specific heat ratio γ), we used the same theoretical curves for deuterium and hydrogen in Fig. 7 and for helium and neon in Fig. 8.

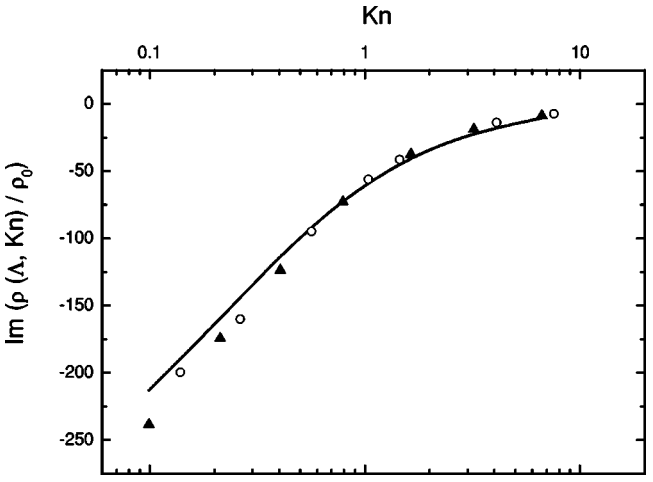


FIG. 8. Comparison of theoretical predictions (solid lines) for the normalized imaginary part of dynamic density at $|\Lambda| = 10^{-1}$ with those resulted from experimental data[29,30] for flows of helium (squares) and neon (asterisks) in flat slits.

Figures 4–9 show that the normalized imaginary part of the dynamic density predicted by the theory (solid lines) correlates well with the distributions that resulted from experiment (symbols) in the Knudsen number range $0 < \mathrm{Kn} < 5$ ($0 < \mathrm{Kn}|\Lambda| < 5$) for all considered cases. This indicates remarkable elasticity of the first-order asymptotic model, which was initially elaborated for small Knudsen numbers. Note that a similar feature was observed in the asymptotic theory[31] developed for heat conductivity of highly porous ceramic materials.

Figures 4–9 also show that the gas rarefaction leads to significant reduction of the dynamic density in low-frequency limit. Parametric computations indicate that the gas rarefaction effect decreases with the dimensionless pa-

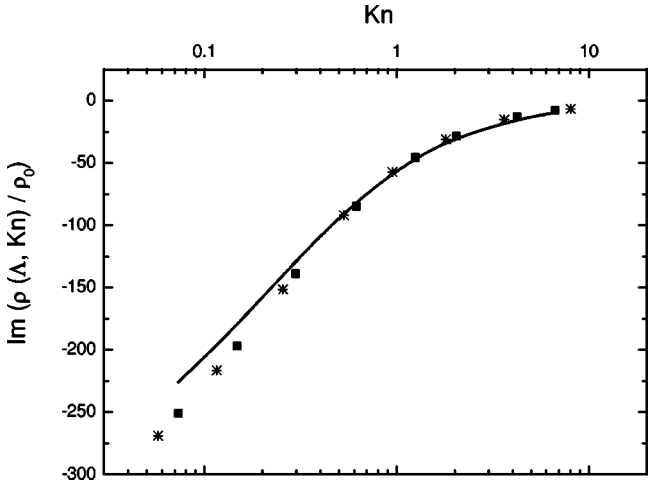


FIG. 7. Comparison of theoretical predictions (solid lines) for the normalized imaginary part of dynamic density at $|\Lambda| = 10^{-1}$ with those resulted from experimental data[29,30] for flows of hydrogen (triangles) and deuterium (circles) in flat slits.

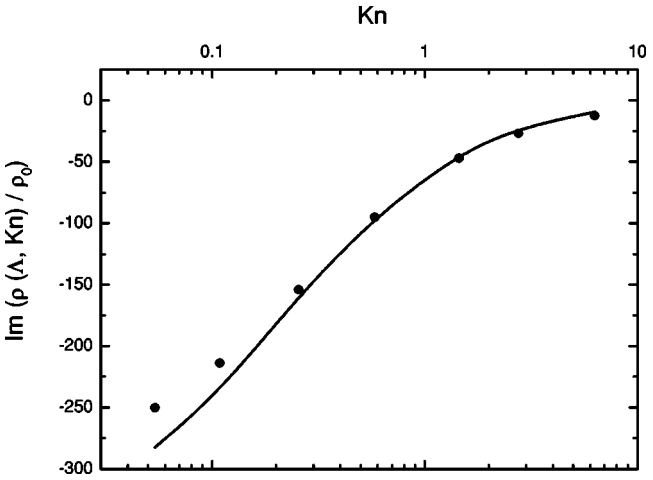


FIG. 9. Comparison of theoretical predictions (solid lines) for the normalized imaginary part of dynamic density at $|\Lambda| = 10^{-1}$ with those resulted from experimental data[29,30] for flows of carbonic gas (rounds) in narrow flat slits.

rameter $|\Lambda|$. This is explained by the reduction of viscous effects on acoustic disturbances of higher frequencies.

Note that the scatter of experimental data presented in Figs. 4–6 increases as the Knudsen number decreases. Reasons for this scatter were not explained in Refs. 26–30. Nevertheless, the rarefaction effect on the dynamic density and other acoustic characteristics diminishes at small Knudsen numbers. Interestingly, the agreement between experiment and theory improves for relatively high Knudsen numbers (Kn$\sim$1) in all considered cases.

V. CONCLUSIONS

In this paper, theoretical analysis of monochromatic acoustic waves propagating in long uniform pores of various cross-sectional shapes (circle, rectangular, triangular, and slit) filled with gas of low density (or in pores of microscopic cross-sectional sizes at normal conditions) was performed. The problem was solved using linearized Navier-Stokes equations with the boundary conditions that resulted from the first-order approximation with respect to small Knudsen number. The obtained results allow prediction of the characteristic impedance Z_w, propagation constant m, and other acoustic parameters of isolated pores and perforated panels in wide ranges of pressures and cross-sectional sizes. For the Knudsen number Kn$=0$, the formulas (29) and (30) along with (31) and (32) coincide with the corresponding expressions for continuum media (see, for example, Ref. 13). In the limiting case of zero frequency, the solutions (5) and (14) describe steady-state isothermal flows (Poiseuille flows) of rarefied gases through long uniform pipes of considered shapes.

The basic theoretical relations were verified by comparisons of the imaginary part of the dynamic density calculated for low frequencies with known experimental data on steady-state flows of rarefied gases through tubes of different cross-sectional shapes. It was shown that theoretical predictions agree well with experiments in the wide range of Knudsen numbers $0 \leqslant \mathrm{Kn} < 5$ that considerably exceeds the formal restriction, $\mathrm{Kn} \ll 1$, of asymptotic theory.

Parametric calculations also showed that the rarefaction effects are stronger in the low-frequency range and substantially decrease at high frequencies. This is due to the corresponding decrease of viscous effects on the dissipation of acoustic energy at higher frequencies.

The results of this paper may be used in the future as a foundation for the development of theoretical and semi-empirical models predicting acoustic characteristics of randomly porous materials in wide ranges of ambient pressure and pore size.

ACKNOWLEDGMENTS

This effort was supported by the Air Force of Scientific Research, Air Force Materials Command under Contracts F499620-99-0005 and F499620-02-C-0024. The U.S government is authorized to reproduce and distribute reprints for government purposes, notwithstanding any copyright notation thereon. The views and conclusions herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed, or implied of the Air Force Office of Scientific Research or the U.S. government.

[1] G. Kirchhoff, "Über den Einfluss der Wärmeleitung in einem Gase auf die Schallbewegung," Poggendorfer Ann. **134**, 177–193 (1868).
[2] J. W. S. Rayleigh, *The Theory of Sound*, (Dover, New York, 1945).
[3] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949).
[4] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. Part I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[5] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. Part 2. High-frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[6] M. E. Delany and E. N. Bazley, "Acoustic properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[7] K. Attenborough, "On the acoustic slow wave in air-filled granular media," J. Acoust. Soc. Am. **81**, 93–102 (1987).
[8] D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[9] J.-F. Allard and Y. Champoux, "Empirical equations for sound propagation in rigid frame porous materials," J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[10] M. R. Stinson, "The propagation of plane sound waves in narrow and wide circular tubes and generalization to uniform tubes of arbitrary cross-sectional shape," J. Acoust. Soc. Am. **89**, 550–558 (1991).
[11] H. S. Roh, W. P. Arnott, J. M. Sabatier, and R. Raspet, "Measurement and calculation of acoustic propagation constants in arrays of small air-filled rectangular tubes," J. Acoust. Soc. Am. **89**, 2617–2624 (1991).
[12] M. R. Stinson and Y. Champoux, "Assignment shape factors for porous materials having simple pore geometries," J. Acoust. Soc. Am. **88**, S121 (1990).
[13] M. R. Stinson and Y. Champoux, "Propagation of sound and the assignment of shape factors of model porous materials having simple pore geometries," J. Acoust. Soc. Am. **91**, 685–695 (1992).
[14] A. Fedorov, A. Shiplyuk, A. Maslov, E. Burov, and N. Malmuth, "Stabilization of a Hypersonic Boundary Layer Using an Ultrasonically Absorptive Coating," J. Fluid Mech. **479**, 99–124 (2003).
[15] A. Fedorov, V. Kozlov, A. Shiplyuk, A. Maslov, A. Sidorenko, E. Burov, and N. Malmuth, "Stability of hypersonic boundary layer on porous wall with regular microstructure," AIAA paper no. 2003-4147.
[16] M. Gad-el-Hak, "Use of continuum and molecular approaches in microfluids," AIAA paper no. 2002-2868.
[17] A. Kundt and E. Warburg, "On Friction and Thermal Conductivity in Rarefies Gases," Philos. Mag. **50**, 53 (1875).
[18] J. C. Maxwell, "On Stresses in Rarefied Gases Arising from Inequalities of Temperature," Philos. Trans. R. Soc. London **170**, 231–256 (1879); reprinted in *The Scientific Papers of J. C. Maxwell* (Dover, New York, 1965).
[19] M. V. Smoluchowski, "Über den Temperatursprung bei Wärmeleitung in Gasen," Akad. Wiss. Wien. **107**, 304 (1898); **108**, 5 (1899).
[20] F. O. Goodman and H. Y. Wachman, *Dynamic of Gas-Surface Scattering* (Academic, New York, 1976).
[21] G. N. Patterson, *Molecular Flow of Gases* (Wiley, New York; Chapman and Hall, London, 1956).
[22] V. P. Shidlovski, *Introduction in Dynamics of Rarefied Gas* (Nauka, Moscow, 1965) (in Russian).
[23] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, 2nd ed. (Cambridge U.P., Cambridge, 1952).
[24] C. Cercignani, *Theory and Application of the Boltzmann Equation* (Scottish Academic, Edinburgh, 1975).
[25] M. Knudsen, "Die Gesetze der Molekularströmung und der inneren Reibungsströmung der Gase durch Rohren," Ann. Phys. **28**, 75–130 (1909).
[26] G. K. Brown, A. DiNardo, G. K. Cheng, and T. K. Sherwood, "The Flow of Gases in Pipes at Low Pressures," J. Appl. Phys. **17**, 802–813 (1946).
[27] M. W. Milligan and H. J. Wilkerson, "Pressure Ratio Effects on Rarefied Flow Through Square Tubes," J. Eng. Ind. **95**, 296–299 (1973).

[28] W. Gaede, ''Die äussere Reibung der Gase,'' Ann. Phys. **41**, 289 (1913).
[29] B. T. Porodnov, P. E. Suetin, and S. F. Borisov, ''Gas flows in flat slit in wide range of Knudsen numbers,'' J. Tech. Phys. **40**, 2383–2391 (1970) (in Russian).
[30] B. T. Porodnov, P. E. Suetin, S. F. Borisov, and V. D. Akinshin, ''Experimental investigation of rarefied gas flow in different channels,'' J. Fluid Mech. **64**, 417–437 (1974).
[31] V. F. Kozlov, ''The theoretical model of heat conductivity in high-porous materials,'' (in Russian), Transactions of First Russian National Conference on Heat Transfer (published by Moscow Power Engineering Institute, Moscow, 1994), vol. X, pt. 2, 3–7.

Multiple scattering by random configurations of circular cylinders: Second-order corrections for the effective wavenumber

C. M. Linton
Department of Mathematical Sciences, Loughborough University, Leicestershire LE11 3TU, United Kingdom

P. A. Martin
Department of Mathematical and Computer Sciences, Colorado School of Mines, Golden, Colorado 80401-1887

(Received 3 September 2004; revised 3 March 2005; accepted 15 March 2005)

A formula for the effective wavenumber in a dilute random array of identical scatterers in two dimensions is derived, based on Lax's quasicrystalline approximation. This formula replaces a widely-used expression due to Twersky, which is shown to be based on an inappropriate choice of pair-correlation function. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1904270]

PACS numbers: 43.20.Fn, 43.20.Bi, 43.20.Hq, 43.28.En [GCG] Pages: 3413–3423

I. INTRODUCTION

Multiple scattering by random arrangements of scatterers is a topic with an extensive literature. See, for example, the recent book by Tsang *et al.* (2001). The modern era dates from the work of Foldy (1945), Lax (1951, 1952), Waterman and Truell (1961), and Twersky (1962). Major applications include wave propagation through suspensions [see, for example McClements *et al.* (1990), Povey (1997), Dukhin and Goetz (2001), and Spelt *et al.* (2001)] and through elastic composites [see Mal and Knopoff (1967), Kim *et al.* (1995), and Kanaun (2000)]. In this paper, we are mainly interested in two-dimensional problems, motivated by the calculation of sound propagation through forests [Embleton (1966); Price *et al.* (1988)]. An important paper on acoustic scattering by arrays of circular cylinders is that of Bose and Mal (1973); see Sec. IV below. For subsequent work, see Varadan *et al.* (1978), Yang and Mal (1994), Bose (1996), Kanaun and Levin (2003), and Kim (2003). For analogous plane-strain elastodynamics, see Varadan *et al.* (1986), Yang and Mal (1994), Bussink *et al.* (1995), and Verbis *et al.* (2001).

A typical problem is the following. The region $x<0$ is filled with a homogeneous compressible fluid of density ρ and sound-speed c. The region $x>0$ contains the same fluid and many scatterers; to fix ideas, we suppose that the scatterers are identical circles (parallel circular cylinders). Then, a time-harmonic plane wave with wavenumber $k=\omega/c$ (ω is the angular frequency) is incident on the scatterers: what is the reflected wave field? This field may be computed exactly for any given configuration (ensemble) of N circles, but the cost increases as N increases. If the computation can be done, it may be repeated for other configurations, and then the average reflected field could be computed (this is the Monte Carlo approach). Instead of doing this, one can try to do some ensemble averaging in order to calculate the average (coherent) field. One result of this is a formula for the *effective wavenumber* K. This can then be used to replace the ''random medium'' occupying $x>0$ by a homogeneous effective medium.

Foldy (1945) began by considering isotropic point scatterers; this is an appropriate model for small sound-soft scatterers. He obtained the formula

$$K^2=k^2-4\mathrm{i}gn_0, \tag{1}$$

where n_0 is the number of circles per unit area and g is the scattering coefficient for an individual scatterer. [In fact, Foldy considered scattering in three dimensions; the two-dimensional formula, Eq. (1), can be found as Eq. (3.20) in Twersky (1962) and Eq. (26) in Aristégui and Angel (2002), for example.] The formula (1) assumes that the scatterers are independent and that n_0 is small. We are interested in calculating the correction to Eq. (1) (a term proportional to n_0^2), and this will require saying more about the distribution of the scatterers; specifically, we shall use pair correlations. Thus, our goal is a formula of the form

$$K^2=k^2+\delta_1 n_0+\delta_2 n_0^2, \tag{2}$$

with computable expressions for δ_1 and δ_2. Moreover, we do not only want to restrict our formula to sound-soft scatterers.

There is some controversy over the proper value for δ_2. In order to state one such formula, we introduce the *far-field pattern* f for scattering by one circular cylinder. Thus, we have $u_{\mathrm{in}}=\exp[\mathrm{i}kr\cos(\theta-\theta_{\mathrm{in}})]$ for the incident plane wave, where (r,θ) are plane polar coordinates ($x=r\cos\theta$, $y=r\sin\theta$) and θ_{in} is the angle of incidence. The scattered waves satisfy

$$u_{\mathrm{sc}}\sim\sqrt{2/(\pi kr)}f(\theta-\theta_{\mathrm{in}})\exp(\mathrm{i}kr-\mathrm{i}\pi/4) \quad \text{as} \quad r\to\infty. \tag{3}$$

Then, Twersky (1962) has given the following formula:

$$K^2=k^2-4\mathrm{i}n_0 f(0)+(2n_0/k)^2\sec^2\theta_{\mathrm{in}}\{[f(\pi-2\theta_{\mathrm{in}})]^2-[f(0)]^2\}. \tag{4}$$

This formula involves θ_{in}, so that it gives a different effective wavenumber for different incident fields.

The three-dimensional version of Eq. (4) is older. For a random collection of identical spheres, it is

$$K^2 = k^2 - 4\pi i(\hat{n}_0/k) f(0) + \delta_2 \hat{n}_0^2$$

with [see Twersky (1962)]

$$\delta_2 = (4\pi^2/k^4)\sec^2\theta_{\rm in}\{[f(\pi - 2\theta_{\rm in})]^2 - [f(0)]^2\}, \qquad (5)$$

where the far-field pattern is now defined by $u_{\rm sc} \sim (ikr)^{-1} e^{ikr} f(\vartheta)$, r and ϑ are spherical polar coordinates, and $\hat{n}_0$ is the number of spheres per unit volume. The same formula but with $\theta_{\rm in} = 0$ (normal incidence) was given by Waterman and Truell (1961). However, it was shown by Lloyd and Berry (1967) that Eq. (5) is incorrect; they obtained

$$\delta_2 = \frac{4\pi^2}{k^4}\left\{ -[f(\pi)]^2 + [f(0)]^2 + \int_0^\pi \frac{1}{\sin(\vartheta/2)} \frac{\mathrm{d}}{\mathrm{d}\vartheta}[f(\vartheta)]^2 \mathrm{d}\vartheta \right\} \qquad (6)$$

(with no dependence on $\theta_{\rm in}$). Lloyd and Berry (1967) used methods (and language) coming from nuclear physics. Thus, in their approach, which they "call the 'resummation method,' a point source of waves is considered to be situated in an infinite medium. The scattering series is then written out completely, giving what Lax has called the 'expanded' representation. In this expanded representation the ensemble average may be taken exactly [but then] the coherent wave does not exist; the series must be resummed in order to obtain any result at all." One purpose of the present paper is to demonstrate that a proper analysis of the semi-infinite two-dimensional model problem (with arbitrary angle of incidence) leads to a formula that is reminiscent of the (three-dimensional) Lloyd–Berry formula; specifically, instead of Eq. (4), we obtain

$$K^2 = k^2 - 4in_0 f(0) + \frac{8n_0^2}{\pi k^2}\int_0^\pi \cot(\theta/2)\frac{\mathrm{d}}{\mathrm{d}\theta}[f(\theta)]^2 \mathrm{d}\theta. \qquad (7)$$

Our analysis does not involve "resumming" series or divergent integrals. It builds on a conventional approach, in the spirit of the papers by Fikioris and Waterman (1964) and by Bose and Mal (1973).

The paper is organized, as follows. Some elementary probability theory is recalled in Sec. II. In particular, the pair-correlation function is introduced; this leads to the notion of "hole correction"—individual cylinders must not be allowed to overlap during the averaging process. In Sec. III, we derive the integral equations of Foldy (isotropic scatterers, no hole correction) and of Lax (isotropic scatterers, hole correction included). Foldy's integral equation can be solved exactly whereas we have been unable to solve Lax's integral equation. Nevertheless, we have developed a rigorous method for extracting an expression for K from these integral equations without actually solving the integral equations themselves. Then, we use the same method in Sec. IV but without the restriction to isotropic scatterers. We start by following Bose and Mal (1973), and use an exact (deterministic) theory for scattering by N circles followed by ensemble averaging. We give a clear derivation of a certain homogeneous infinite linear system of algebraic equations, obtained previously by Bose and Mal (1973) for the case of normal incidence; the system does not depend on $\theta_{\rm in}$ and the existence of a nontrivial solution determines K. We solve the system for small n_0, and obtain Eq. (7). We also show that Eq. (4) is obtained if the hole correction is not done correctly. Concluding remarks are given in Sec. V.

II. SOME PROBABILITY THEORY

In this section, we give a very brief summary of the probability theory needed. For more information, see Foldy (1945), Lax (1951), Aristégui and Angel (2002) or Chap. 14 of Ishimaru (1978).

Suppose we have N scatterers located at the points $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$; denote the configuration of points by $\Lambda_N = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$. Then, the ensemble (or configurational) average of any quantity $F(\mathbf{r}|\Lambda_N)$ is defined by

$$\langle F(\mathbf{r})\rangle = \int \cdots \int p(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) F(\mathbf{r}|\Lambda_N) \mathrm{d}V_1 \cdots \mathrm{d}V_N, \qquad (8)$$

where the integration is over N copies of the volume B_N containing N scatterers. Here, $p(\mathbf{r}_1, \ldots, \mathbf{r}_N)\mathrm{d}V_1 \mathrm{d}V_2 \cdots \mathrm{d}V_N$ is the probability of finding the scatterers in a configuration in which the first scatterer is in the volume element $\mathrm{d}V_1$ about $\mathbf{r}_1$, the second scatterer is in the volume element $\mathrm{d}V_2$ about $\mathbf{r}_2$, and so on, up to $\mathbf{r}_N$. The joint probability distribution $p(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is normalized so that $\langle 1\rangle = 1$. Similarly, the average of $F(\mathbf{r}|\Lambda_N)$ over all configurations for which the first scatterer is fixed at $\mathbf{r}_1$ is given by

$$\langle F(\mathbf{r})\rangle_1 = \int \cdots \int p(\mathbf{r}_2, \ldots, \mathbf{r}_N|\mathbf{r}_1) F(\mathbf{r}|\Lambda_N) \mathrm{d}V_2 \cdots \mathrm{d}V_N, \qquad (9)$$

where $p(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = p(\mathbf{r}_1) p(\mathbf{r}_2, \ldots, \mathbf{r}_N|\mathbf{r}_1)$ defines the conditional probability $p(\mathbf{r}_2, \ldots, \mathbf{r}_N|\mathbf{r}_1)$. If two scatterers are fixed, say the first and the second, we can define

$$\langle F(\mathbf{r})\rangle_{12} = \int \cdots \int p(\mathbf{r}_3, \ldots, \mathbf{r}_N|\mathbf{r}_1, \mathbf{r}_2) F(\mathbf{r}|\Lambda_N) \mathrm{d}V_3 \cdots \mathrm{d}V_N, \qquad (10)$$

where $p(\mathbf{r}_2, \ldots, \mathbf{r}_N|\mathbf{r}_1) = p(\mathbf{r}_2|\mathbf{r}_1) p(\mathbf{r}_3, \ldots, \mathbf{r}_N|\mathbf{r}_1, \mathbf{r}_2)$.

Now, as each of the N scatterers is equally likely to occupy $\mathrm{d}V_1$, the density of scatterers at $\mathbf{r}_1$ is $Np(\mathbf{r}_1) = n_0$, the (constant) number of scatterers per unit volume. Thus

$$p(\mathbf{r}) = n_0/N = |B_N|^{-1}, \qquad (11)$$

where $|B_N|$ is the volume of B_N. Also, as $p(\mathbf{r}_1, \mathbf{r}_2) = p(\mathbf{r}_1) p(\mathbf{r}_2|\mathbf{r}_1)$, we obtain

$$\int\int p(\mathbf{r}_2|\mathbf{r}_1)\mathrm{d}V_1 \mathrm{d}V_2 = \frac{N}{n_0} = |B_N|. \qquad (12)$$

We have to specify $p(\mathbf{r}_2|\mathbf{r}_1)$, consistent with Eq. (12). Also, we want to ensure that scatterers do not overlap. For circular cylinders of radius a, a simple choice is $p(\mathbf{r}_2|\mathbf{r}_1) = p_0 H(R_{12} - b)$ with $b \geqslant 2a$, where $H(x)$ is the Heaviside unit function, $R_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$ and p_0 is a constant determined by Eq. (12). Thus,

$$p_0 = \{|B_N| - \pi b^2\}^{-1} \simeq n_0/N, \qquad (13)$$

assuming that $b^2 n_0/N \ll 1$. [The equality in Eq. (13) assumes that the "hole" at $\mathbf{r}_1$ of radius b does not cut the boundary of B_N. Evidently, taking this possibility into account would not change the approximation $p_0 \simeq n_0/N$.] Hence, the simplest sensible choice for the pair-correlation function is

$$p(\mathbf{r}_2|\mathbf{r}_1) = \begin{cases} 0, & R_{12} < b, \\ n_0/N, & R_{12} > b. \end{cases} \tag{14}$$

This simple choice will be used for most of our analysis. More generally, we could use

$$p(\mathbf{r}_2|\mathbf{r}_1) = \begin{cases} 0, & R_{12} < b, \\ (n_0/N)[1+\chi(R_{12};n_0)], & R_{12} > b, \end{cases} \tag{15}$$

where the function χ is to be chosen, subject to some constraints. The effect of using Eq. (15) instead of Eq. (14) is calculated in Sec. IV D. One could also consider functions χ that depend on $\mathbf{r}_1 - \mathbf{r}_2$ (instead of just $|\mathbf{r}_1 - \mathbf{r}_2|$); such possibilities are discussed in Twersky (1978) and Siqueira *et al.* (1995).

III. FOLDY–LAX THEORY: ISOTROPIC SCATTERERS

Foldy's theory begins with a simplified deterministic model for scattering by N identical scatterers, each of which is supposed to scatter isotropically. Thus, the total field is assumed to be given by the incident field plus a point source at each scattering center, $\mathbf{r}_j$:

$$u(\mathbf{r}|\Lambda_N) = u_{\text{in}}(\mathbf{r}) + g\sum_{j=1}^{N} u_{\text{ex}}(\mathbf{r}_j;\mathbf{r}_j|\Lambda_N) H_0(k|\mathbf{r}-\mathbf{r}_j|). \tag{16}$$

Here, $H_n(w) \equiv H_n^{(1)}(w)$ is a Hankel function, g is the (assumed known) scattering coefficient, and the exciting field u_{ex} is given by

$$u_{\text{ex}}(\mathbf{r};\mathbf{r}_n|\Lambda_N) = u_{\text{in}}(\mathbf{r}) + g\sum_{\substack{j=1 \\ j\neq n}}^{N} u_{\text{ex}}(\mathbf{r}_j;\mathbf{r}_j|\Lambda_N) H_0(k|\mathbf{r}-\mathbf{r}_j|). \tag{17}$$

The second term in Eq. (17) is the field near the cylinder at $\mathbf{r}_n$ due to scattering by all the other cylinders. The N numbers $u_{\text{ex}}(\mathbf{r}_j;\mathbf{r}_j|\Lambda_N)$ $(j=1,2,...,N)$ required in Eq. (16) are to be determined by solving the linear system obtained by evaluating Eq. (17) at $\mathbf{r}=\mathbf{r}_n$; direct numerical solutions of this system have been given by Fikioris (1966) and by Groenenboom and Snieder (1995).

Let us try to compute the ensemble average of u, using Eqs. (16) and (8). The result is

$$\langle u(\mathbf{r})\rangle = u_{\text{in}}(\mathbf{r}) + gn_0 \int_{B_N} \langle u_{\text{ex}}(\mathbf{r}_1)\rangle_1 H_0(k|\mathbf{r}-\mathbf{r}_1|)\,\mathrm{d}V_1, \tag{18}$$

where we have used Eqs. (9) and (11), and the indistinguishability of the scatterers. For $\langle u_{\text{ex}}(\mathbf{r}_1)\rangle_1$ [which is given explicitly by Eq. (9) in which $u_{\text{ex}}(\mathbf{r}_1;\mathbf{r}_1|\Lambda_N)$ is substituted for $F(\mathbf{r}|\Lambda_N)$], we obtain

$$\langle u_{\text{ex}}(\mathbf{r})\rangle_1 = u_{\text{in}}(\mathbf{r}) + g(N-1)\int_{B_N} p(\mathbf{r}_2|\mathbf{r}_1) \times \langle u_{\text{ex}}(\mathbf{r}_2)\rangle_{12} H_0(k|\mathbf{r}-\mathbf{r}_2|)\,\mathrm{d}V_2, \tag{19}$$

where we have used Eqs. (10) and (17). Equations (18) and (19) are the first two in a hierarchy, involving more and more complicated information on the statistics of the scatterer distribution. In practice, the hierarchy is broken using an additional assumption. At the lowest level, we have Foldy's assumption,

$$\langle u_{\text{ex}}(\mathbf{r})\rangle_1 \simeq \langle u(\mathbf{r})\rangle, \tag{20}$$

at least in the neighborhood of $\mathbf{r}_1$. When this is used in Eq. (18), we obtain

$$\langle u(\mathbf{r})\rangle = u_{\text{in}}(\mathbf{r}) + gn_0 \int_{B_N} \langle u(\mathbf{r}_1)\rangle H_0(k|\mathbf{r}-\mathbf{r}_1|)\,\mathrm{d}V_1, \quad \mathbf{r}\in B_N. \tag{21}$$

We call this *Foldy's integral equation* for $\langle u\rangle$. The integral on the right-hand side is an acoustic volume potential. Hence, an application of $(\nabla^2 + k^2)$ to Eq. (21) eliminates the incident field and shows that $(\nabla^2 + K^2)\langle u\rangle = 0$ in B_N, where K^2 is given by Foldy's formula, Eq. (1).

At the next level, we have the Lax (1952) quasicrystalline assumption (QCA),

$$\langle u_{\text{ex}}(\mathbf{r})\rangle_{12} \simeq \langle u_{\text{ex}}(\mathbf{r})\rangle_2. \tag{22}$$

When this is used in Eq. (19) evaluated at $\mathbf{r}=\mathbf{r}_1$, we obtain

$$v(\mathbf{r}) = u_{\text{in}}(\mathbf{r}) + g(N-1)\int_{B_N} p(\mathbf{r}_1|\mathbf{r}) v(\mathbf{r}_1) H_0(k|\mathbf{r}-\mathbf{r}_1|)\,\mathrm{d}V_1, \quad \mathbf{r}\in B_N, \tag{23}$$

where $v(\mathbf{r}) = \langle u_{\text{ex}}(\mathbf{r})\rangle_1$. We call this *Lax's integral equation*.

In what follows, we let $N\to\infty$ so that $B_N \to B_\infty$, a semi-infinite region, $x>0$.

A. Foldy's integral equation: Exact treatment

Consider a plane wave at oblique incidence, so that

$$u_{\text{in}} = \mathrm{e}^{\mathrm{i}(\alpha x + \beta y)} \quad \text{with} \quad \alpha = k\cos\theta_{\text{in}} \quad \text{and} \quad \beta = k\sin\theta_{\text{in}}. \tag{24}$$

For a semi-infinite domain B_∞, Foldy's integral equation, Eq. (21), becomes

$$\langle u(x,y)\rangle = \mathrm{e}^{\mathrm{i}(\alpha x+\beta y)} + gn_0 \int_0^\infty \int_{-\infty}^{\infty} \langle u(x_1, y+Y)\rangle \times H_0(k\rho_1)\,\mathrm{d}Y\,\mathrm{d}x_1, \quad \begin{matrix} x>0, \\ -\infty<y<\infty, \end{matrix}$$

where $\rho_1 = \sqrt{(x-x_1)^2 + Y^2}$. This equation can be solved exactly. Thus, writing

$$\langle u(x,y)\rangle = U(x)\mathrm{e}^{\mathrm{i}\mu y}, \quad x>0, \quad -\infty<y<\infty, \tag{25}$$

we obtain

$$U(x)=e^{i\alpha x}e^{i(\beta-\mu)y}+gn_0\int_0^\infty\int_{-\infty}^{\infty}U(x_1)$$

$$\times H_0(k\rho_1)e^{i\mu Y}dYdx_1, \quad \begin{matrix} x>0, \\ -\infty<y<\infty. \end{matrix} \tag{26}$$

Hence, for a solution in the form Eq. (25), we must have $\mu=\beta=k\sin\theta_{\rm in}$.

Now,

$$\int_{-\infty}^{\infty}H_0(k\rho_1)e^{i\beta Y}dY=\frac{2}{\alpha}e^{i\alpha|x-x_1|}, \tag{27}$$

where $\alpha=\sqrt{k^2-\beta^2}=k\cos\theta_{\rm in}$. Thus, we see that U solves

$$U(x)=e^{i\alpha x}+\frac{2gn_0}{\alpha}\int_0^\infty U(x_1)e^{i\alpha|x-x_1|}dx_1, \quad x>0. \tag{28}$$

Now, put $U(x)=U_0e^{i\lambda x}$, so that Eq. (25) gives

$$\langle u(x,y)\rangle=U_0e^{i(\lambda x+\beta y)}, \quad x>0, \quad -\infty<y<\infty, \tag{29}$$

and Eq. (28) gives

$$U_0e^{i\lambda x}-e^{i\alpha x}=\frac{2gn_0U_0}{i\alpha}\left(\frac{2\alpha e^{i\lambda x}}{\lambda^2-\alpha^2}-\frac{e^{i\alpha x}}{\lambda-\alpha}\right),$$

where we have assumed that $\mathrm{Im}\,\lambda>0$. If we compare the coefficients of $e^{i\lambda x}$, we see that U_0 cancels, leaving

$$\lambda^2-\alpha^2=-4ign_0, \tag{30}$$

which determines λ. Then, the coefficients of $e^{i\alpha x}$ give $U_0=2\alpha/(\lambda+\alpha)$. A similar method can be used to find $\langle u\rangle$ when B_∞ is a slab of finite thickness, $0<x<h$, say; see Aristégui and Angel (2002).

From Eq. (29), it is natural to write

$$\lambda=K\cos\varphi \quad \text{and} \quad \beta=K\sin\varphi=k\sin\theta_{\rm in}. \tag{31}$$

These define the effective wave number K; the last equality is recognized as Snell's law, even though K and φ are complex, with $\mathrm{Im}\,K>0$. Hence, we see that

$$\lambda^2-\alpha^2=K^2-k^2, \tag{32}$$

and so Eq. (30) reduces to Foldy's formula, Eq. (1).

B. Foldy's integral equation: Approximate treatment

We have seen that Foldy's integral equation can be solved exactly, and that the solution process has two parts: first find λ (and hence the effective wavenumber) and then find U_0. In fact, λ can be found without finding the complete solution; the reason for pursuing this is that we cannot usually find exact solutions. Thus, consider Eq. (28), and suppose that

$$U(x)=U_0e^{i\lambda x} \quad \text{for } x>\ell,$$

where U_0, λ, and ℓ are unknown. To proceed, we need say nothing about the solution U in the "boundary layer" $0<x<\ell$. Now, evaluate the integral equation for $x>\ell$; we find that

$$U_0e^{i\lambda x}-e^{i\alpha x}$$

$$=\frac{2gn_0}{\alpha}e^{i\alpha x}\int_0^\ell U(t)e^{-i\alpha t}dt+\frac{2gn_0}{\alpha}$$

$$\times\int_\ell^\infty U(t)e^{i\alpha|x-t|}dt=\mathcal{A}e^{i\lambda x}+\mathcal{B}e^{i\alpha x} \quad \text{for } x>\ell,$$

where $\mathcal{A}=-4igU_0/(\lambda^2-\alpha^2)$ and

$$\mathcal{B}=\frac{2gn_0}{\alpha}\int_0^\ell U(t)e^{-i\alpha t}dt+\frac{2igU_0}{\alpha(\lambda-\alpha)}e^{i(\lambda-\alpha)\ell}.$$

Then, setting $U_0=\mathcal{A}$ gives Eq. (30) again, without knowing the solution U everywhere. This basic method will be used again below.

C. Lax's integral equation

Using the approximation $p(\mathbf{r}_1|\mathbf{r})=(n_0/N)H(R_1-b)$ in Lax's integral equation, Eq. (23) gives

$$v(\mathbf{r})=u_{\rm in}(\mathbf{r})+gn_0\frac{N-1}{N}\int_{B_N^b(\mathbf{r})}v(\mathbf{r}_1)H_0(kR_1)d\mathbf{r}_1,$$

$$\mathbf{r}\in B_N, \tag{33}$$

where $B_N^b(\mathbf{r})=\{\mathbf{r}_1\in B_N: R_1=|\mathbf{r}-\mathbf{r}_1|>b\}$, which is B_N with a (possibly incomplete) disk excluded.

Let $N\to\infty$ and take an incident plane wave, Eq. (24), giving

$$v(x,y)=e^{i(\alpha x+\beta y)}+gn_0\int_{x_1>0,\rho_1>b}v(x_1,y+Y)$$

$$\times H_0(k\rho_1)dYdx_1, \quad \begin{matrix} x>0, \\ -\infty<y<\infty. \end{matrix}$$

As in Sec. III A, we write

$$v(x,y)=V(x)e^{i\beta y}, \quad x>0, \quad -\infty<y<\infty, \tag{34}$$

giving

$$V(x)=e^{i\alpha x}+gn_0\int_{x_1>0,\rho_1>b}V(x_1)H_0(k\rho_1)$$

$$\times e^{i\beta Y}dYdx_1, \quad x>0. \tag{35}$$

Then, using Eq. (27), we see that V solves

$$V(x)=e^{i\alpha x}+gn_0\int_0^\infty V(x_1)L(x-x_1)dx_1, \quad x>0, \tag{36}$$

where the kernel, $L(x-x_1)$, is given by

$$L(X)=\frac{2}{\alpha}e^{i\alpha|X|}-2\int_0^{c(X)}H_0(k\sqrt{X^2+Y^2})e^{i\beta Y}dY \tag{37}$$

with $c(X)=\sqrt{b^2-X^2}H(b-|X|)$; here, we have written the integral over Y in Eq. (35) as an integral over all Y minus an integral through the disk, if necessary.

We have been unable to solve Eq. (36) exactly (even though it is an integral equation of Wiener–Hopf-type). However, the approximate method described in Sec. III B can be used. Thus, let us suppose that

$$V(x)=V_0 e^{i\lambda x} \quad \text{for } x>\ell, \tag{38}$$

where V_0, λ, and ℓ are unknown. Then, consider Eq. (36) for $x>\ell+b$, so that the interval $|x-x_1|<b$ is entirely within the range $x_1>\ell$. Making use of Eq. (37), Eq. (36) gives

$$\frac{V_0 e^{i\lambda x}-e^{i\alpha x}}{gn_0} = \frac{2}{\alpha}e^{i\alpha x}\int_0^{\ell} V(t)e^{-i\alpha t}dt+\frac{2}{\alpha}\int_{\ell}^{\infty} V(t)e^{i\alpha|x-t|}dt - 2\int_{x-b}^{x+b} V(t)\int_0^{c(x-t)} H_0(k\sqrt{(x-t)^2+Y^2})e^{i\beta Y}dY\,dt \tag{39}$$

for $x>\ell+b$. Equation (38) can be used in the second and third integrals. The second integral is elementary, and has the value

$$\frac{2iV_0}{\alpha(\lambda-\alpha)}e^{i(\lambda-\alpha)\ell}e^{i\alpha x}-\frac{4iV_0}{\lambda^2-\alpha^2}e^{i\lambda x}.$$

The third integral becomes

$$-2V_0\int_{-b}^{b} e^{i\lambda(x+\xi)}\int_0^{\sqrt{b^2-\xi^2}} H_0(k\sqrt{\xi^2+Y^2})e^{i\beta Y}dY\,d\xi = -V_0 e^{i\lambda x}\int_0^{2\pi}\int_0^{b} e^{iKr\cos(\theta-\varphi)}H_0(kr)r\,dr\,d\theta = -2\pi V_0 e^{i\lambda x}\int_0^{b} J_0(Kr)H_0(kr)r\,dr = V_0 e^{i\lambda x}\left\{\frac{4i}{K^2-k^2}-\frac{2\pi\mathcal{N}_0(Kb)}{K^2-k^2}\right\},$$

where $\mathcal{N}_0(Kb)=KbH_0(kb)J_1(Kb)-kbH_1(kb)J_0(Kb)$. Using these results in Eq. (39), noting Eq. (32), we obtain

$$V_0 e^{i\lambda x}-e^{i\alpha x}=\mathcal{A}e^{i\lambda x}+\mathcal{B}e^{i\alpha x} \quad \text{for } x>\ell+b,$$

where

$$\mathcal{A}=\frac{2\pi g n_0 V_0}{k^2-K^2}\mathcal{N}_0(Kb),$$

$$\mathcal{B}=\frac{2gn_0}{\alpha}\int_0^{\ell} V(t)e^{-i\alpha t}dt+\frac{2ign_0V_0}{\alpha(\lambda-\alpha)}e^{i(\lambda-\alpha)\ell}.$$

For a solution, we must have $\mathcal{A}=V_0$, and so

$$K^2=k^2-2\pi g n_0\mathcal{N}_0(Kb), \tag{40}$$

which is a nonlinear equation for K. Notice that this equation does not depend on the angle of incidence, θ_{in}.

We have $\mathcal{N}_0(Kb)\to 2i/\pi$ as $b\to 0$ so that, in this limit, we recover Foldy's formula for the effective wavenumber, Eq. (1).

Let us solve Eq. (40) for small n_0. (Alternatively, we could use the dimensionless area fraction $\pi a^2 n_0$.) Begin by writing

$$K^2=k^2+\delta_1 n_0+\delta_2 n_0^2+\cdots, \tag{41}$$

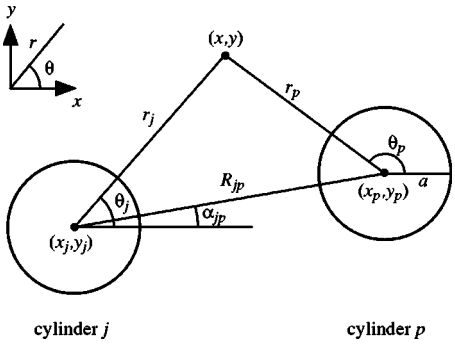


FIG. 1. A view of two typical cylinders.

where δ_1 and δ_2 are to be found; for δ_1, we expect to obtain the result given by Eq. (1). It follows from Eq. (41) that $K=k+\frac{1}{2}\delta_1 n_0/k+O(n_0^2)$ and then

$$\mathcal{N}_0(Kb)=\mathcal{N}_0(kb)+(Kb-kb)\mathcal{N}_0'(kb)+\cdots = 2i/\pi+\tfrac{1}{2}b^2\delta_1 d_0(kb)n_0+O(n_0^2),$$

where $d_0(x)=J_0(x)H_0(x)+J_1(x)H_1(x)$. When this approximation for $\mathcal{N}_0(Kb)$ is used in Eq. (40), we obtain

$$K^2=k^2-4ign_0-\pi b^2 g\delta_1 d_0(kb)n_0^2.$$

Comparison of this formula with Eq. (41) gives $\delta_1=-4ig$ (as expected) and $\delta_2=4\pi i(gb)^2 d_0(kb)$, so that we obtain the approximation

$$K^2=k^2-4ign_0+4\pi i(gbn_0)^2 d_0(kb). \tag{42}$$

Note that the second-order term in Eq. (42) vanishes in the limit $kb\to 0$.

IV. FINITE-SIZE EFFECTS

The theory described above relies on the assumption of isotropy. Here, we use a more complete theory. We start with an exact theory (due to Záviška) for acoustic scattering by N identical circular cylinders of radius a; for details and references, see p. 173 of Linton and McIver (2001). The cylinders can be soft, hard or penetrable. Then (in Sec. IV B), we form averaged equations, and we invoke the QCA. This leads to an infinite homogeneous system of linear algebraic equations from which the effective wave number, K, is to be determined; the equations are independent of the angle of incidence. An approximate solution for K is found in Sec. IV C, correct to $O(n_0^2)$. In Sec. IV D, it is shown that this approximation does not depend on the choice of the function $\chi(r;n_0)$, appearing in the pair-correlation function, Eq. (15). In Sec. IV E, it is shown how Twersky's formula for K can be derived, using an unreasonable choice for the pair-correlation function.

A. A finite array of identical circular cylinders: Exact theory

We use polar coordinates (r,θ) centered at the origin and (r_j,θ_j), centered at $\mathbf{r}_j=(x_j,y_j)$, the center of the jth cylinder. The various parameters relating to the relative positions of the cylinders are shown in Fig. 1.

Exterior to the cylinders the pressure field is u, where

$$\nabla^2 u + k^2 u = 0. \tag{43}$$

In the interior of cylinder j, the field is u_j, where

$$\nabla^2 u_j + \kappa^2 u_j = 0. \tag{44}$$

A plane wave, given by Eq. (24), is incident on the cylinders. A phase factor for each cylinder, I_j, is defined by

$$I_j = \mathrm{e}^{\mathrm{i}(\alpha x_j + \beta y_j)} \tag{45}$$

and then we can write

$$u_{\mathrm{in}} = I_j \mathrm{e}^{\mathrm{i}kr_j \cos(\theta_j - \theta_{\mathrm{in}})} = I_j \sum_{n=-\infty}^{\infty} \mathrm{e}^{\mathrm{i}n(\pi/2 - \theta_j + \theta_{\mathrm{in}})} J_n(kr_j). \tag{46}$$

We seek a solution to Eqs. (43) and (44) in the form

$$u = u_{\mathrm{in}} + \sum_{j=1}^{N} \sum_{n=-\infty}^{\infty} A_n^j Z_n H_n(kr_j) \mathrm{e}^{\mathrm{i}n\theta_j}, \tag{47}$$

$$u_j = \sum_{n=-\infty}^{\infty} B_n^j J_n(\kappa r_j) \mathrm{e}^{\mathrm{i}n\theta_j}, \tag{48}$$

for some set of unknown complex coefficients A_n^j and B_n^j. The factor

$$Z_n = \frac{qJ_n'(ka)J_n(\kappa a) - J_n(ka)J_n'(\kappa a)}{qH_n'(ka)J_n(\kappa a) - H_n(ka)J_n'(\kappa a)} = Z_{-n} \tag{49}$$

has been introduced for later convenience. Here $\kappa = \omega/\tilde{c}$ and $q = \tilde{\rho}\tilde{c}/(\rho c)$, where $\tilde{\rho}$ and $\tilde{c}$ are the density and sound speed, respectively, inside the cylinders. Note that we recover the sound-soft results in the limit $q \to 0$, whereas the limit $q \to \infty$ gives the sound-hard results. The boundary conditions on the cylinders are

$$u = u_s, \quad \frac{1}{\rho}\frac{\partial u}{\partial r_s} = \frac{1}{\tilde{\rho}}\frac{\partial u_s}{\partial r_s} \quad \text{on } r_s = a, \quad s = 1,...,N. \tag{50}$$

Using Graf's addition theorem for Bessel functions, it can be shown that provided $r_s < R_{js}$ for all j, we can write the field exterior to cylinder s as

$$\begin{aligned} u(r_s, \theta_s) &= \sum_{n=-\infty}^{\infty} (I_s J_n(kr_s) \mathrm{e}^{\mathrm{i}n(\pi/2 - \theta_s + \theta_{\mathrm{in}})} + A_n^s Z_n H_n(kr_s) \mathrm{e}^{\mathrm{i}n\theta_s}) \\ &\quad + \sum_{\substack{j=1 \\ j \neq s}}^{N} \sum_{n=-\infty}^{\infty} A_n^j Z_n \sum_{m=-\infty}^{\infty} J_m(kr_s) H_{n-m}(kR_{js}) \\ &\quad \times \mathrm{e}^{\mathrm{i}m\theta_s} \mathrm{e}^{\mathrm{i}(n-m)\alpha_{js}}. \end{aligned} \tag{51}$$

The geometrical restriction implies that this expression is only valid if the point (r_s, θ_s) is closer to the center of cylinder s than the centers of any of the other cylinders. This is certainly true on the surface of cylinder s and so Eq. (51) can be used to apply the body boundary conditions which leads, after using the orthogonality of the functions $\exp(\mathrm{i}m\theta_s)$, $m \in \mathbb{Z}$, and eliminating the coefficients B_n^j, to the system of equations

$$\begin{aligned} &A_m^s + \sum_{\substack{j=1 \\ j \neq s}}^{N} \sum_{n=-\infty}^{\infty} A_n^j Z_n \mathrm{e}^{\mathrm{i}(n-m)\alpha_{js}} H_{n-m}(kR_{js}) \\ &\quad = -I_s \mathrm{e}^{\mathrm{i}m(\pi/2 - \theta_{\mathrm{in}})}, \quad \begin{matrix} s = 1,2,...,N, \\ m \in \mathbb{Z}. \end{matrix} \end{aligned} \tag{52}$$

Note that the quantities q, κ, and a only enter the equations through the terms Z_n.

For a single cylinder the solution is immediate: $A_m^1 = -\mathrm{i}^m I_1 \exp(-\mathrm{i}m\theta_{\mathrm{in}})$ and then the far-field pattern, defined by Eq. (3), is given by

$$f(\theta) = -\sum_{n=-\infty}^{\infty} Z_n \mathrm{e}^{\mathrm{i}n\theta}. \tag{53}$$

B. Arrays of circular cylinders: Averaged equations

The above analysis applies to a specific configuration of scatterers. Now we follow Bose and Mal (1973) and take ensemble averages. Specifically, setting $s = 1$ in Eq. (52) and then taking the conditional average, using Eq. (14), we get

$$\begin{aligned} &\langle A_m^1 \rangle_1 + n_0 \frac{N-1}{N} \sum_{n=-\infty}^{\infty} Z_n \int_{B_N : R_{12} > b} H_{n-m}(kR_{21}) \\ &\quad \times \mathrm{e}^{\mathrm{i}(n-m)\alpha_{21}} \langle A_n^2 \rangle_{12} \mathrm{d}V_2 \\ &\quad = -I_1 \mathrm{e}^{\mathrm{i}m(\pi/2 - \theta_{\mathrm{in}})}, \quad m \in \mathbb{Z}. \end{aligned} \tag{54}$$

Now we let $N \to \infty$ so that B_N becomes the half-space $x > 0$, and invoke Lax's QCA, Eq. (22). This implies that

$$\langle A_m^2 \rangle_{12} = \langle A_m^2 \rangle_2. \tag{55}$$

We seek a solution to Eq. (54) in the form

$$\langle A_m^s \rangle_s = \mathrm{i}^m \mathrm{e}^{\mathrm{i}\beta y_s} \Phi_m(x_s) \tag{56}$$

so that

$$\begin{aligned} &\Phi_m(x_1) + n_0 \sum_{n=-\infty}^{\infty} Z_n (-\mathrm{i})^{n-m} \int_{x_2 > 0, R_{12} > b} \psi_{n-m}(x_{21}, y_{21}) \\ &\quad \times \mathrm{e}^{\mathrm{i}\beta y_{21}} \Phi_n(x_2) \mathrm{d}x_2 \mathrm{d}y_2 \\ &\quad = -\mathrm{e}^{-\mathrm{i}m\theta_{\mathrm{in}}} \mathrm{e}^{\mathrm{i}\alpha x_1}, \quad m \in \mathbb{Z}, \end{aligned} \tag{57}$$

where we have written $x_{21} = x_2 - x_1$ and $y_{21} = y_2 - y_1$, used $\alpha_{21} = \alpha_{12} - \pi$, and defined $\psi_n(X,Y) = H_n(kR)\mathrm{e}^{\mathrm{i}n\Theta}$ with $X = R\cos\Theta$ and $Y = R\sin\Theta$.

Proceeding as before, suppose that for sufficiently large x (say $x > \ell$) we can write

$$\Phi_m(x) = F_m \mathrm{e}^{-\mathrm{i}m\varphi} \mathrm{e}^{\mathrm{i}\lambda x}, \tag{58}$$

where λ and φ are defined by Eq. (31). We assume that $\mathrm{Im}\,\lambda > 0$ so that $\Phi_m \to 0$ as $x \to \infty$. Then if $x_1 > \ell + b$, Eq. (57) becomes

$$\begin{aligned} &F_m \mathrm{e}^{-\mathrm{i}m\varphi} \mathrm{e}^{\mathrm{i}\lambda x_1} + n_0 \sum_{n=-\infty}^{\infty} Z_n (-\mathrm{i})^{n-m} \\ &\quad \times \left\{ \int_0^{\ell} \Phi_n(x_2) L_{n-m}(x_{21}) \mathrm{d}x_2 + F_n \mathrm{e}^{-\mathrm{i}n\varphi} \mathrm{e}^{\mathrm{i}\lambda x_1} M_{n-m} \right\} \\ &\quad = -\mathrm{e}^{-\mathrm{i}m\theta_{\mathrm{in}}} \mathrm{e}^{\mathrm{i}\alpha x_1}, \quad m \in \mathbb{Z}, \end{aligned} \tag{59}$$

where

$$L_n(X)=\int_{-\infty}^{\infty}\psi_n(X,Y)\mathrm{e}^{\mathrm{i}\beta Y}\mathrm{d}Y, \tag{60}$$

$$M_n=\int_{x_2>\ell,R_{12}>b}\psi_n(x_{21},y_{21})\Psi(x_{21},y_{21})\mathrm{d}x_2\mathrm{d}y_2, \tag{61}$$

$$\Psi(X,Y)=\mathrm{e}^{\mathrm{i}(\lambda X+\beta Y)}=\mathrm{e}^{\mathrm{i}KR\cos(\Theta-\varphi)}, \tag{62}$$

and we have used Eq. (31). Next, we shall evaluate L_n and M_n; note that we have $x_2<\ell<x_1$ in Eq. (59) so that $x_{21}<0$.

Consider the integral $L_n(X)$ for $X<0$. From Eq. (27), we have

$$L_0(X)=(2/\alpha)\mathrm{e}^{-\mathrm{i}\alpha X}\quad\text{and}\quad L_0'=-\mathrm{i}\alpha L_0. \tag{63}$$

For L_n with $n>0$, we use the fact that

$$L_n(X)=\int_{-\infty}^{\infty}-\frac{1}{k}\mathrm{e}^{\mathrm{i}\beta Y}\left(\frac{\partial}{\partial X}+\mathrm{i}\frac{\partial}{\partial Y}\right)\psi_{n-1}(X,Y)\mathrm{d}Y. \tag{64}$$

Then, as the $\partial/\partial X$ can be taken outside the integral and the $\partial/\partial Y$ can be removed using an integration by parts, we have $kL_n=-L_{n-1}'-\beta L_{n-1}$, which expresses L_n in terms of L_{n-1} and L_{n-1}'. It follows from Eq. (63) that

$$L_n=\frac{2(\mathrm{i}\alpha-\beta)^n}{\alpha k^n}\mathrm{e}^{-\mathrm{i}\alpha X}=\frac{2\mathrm{i}^n}{\alpha}\mathrm{e}^{\mathrm{i}n\theta_{\mathrm{in}}}\mathrm{e}^{-\mathrm{i}\alpha X}. \tag{65}$$

This formula also holds for $n<0$. Hence, for $x_1>x_2$,

$$L_n(x_2-x_1)=(2/\alpha)\mathrm{i}^n\mathrm{e}^{\mathrm{i}n\theta_{\mathrm{in}}}\mathrm{e}^{\mathrm{i}\alpha(x_1-x_2)}. \tag{66}$$

The double integral M_n can be evaluated using Green's theorem as follows. We have $\psi_n\nabla^2\Psi-\Psi\nabla^2\psi_n=(k^2-K^2)\psi_n\Psi$. It follows that

$$M_m=\frac{1}{k^2-K^2}\int_{\partial B}\left[\psi_m\frac{\partial\Psi}{\partial n}-\Psi\frac{\partial\psi_m}{\partial n}\right]\mathrm{d}s_2,$$

where ∂B consists of two parts, the line $x_2=\ell$ and the circle $R_{12}=b$. Now, on $x_2=\ell$, $\partial/\partial n=-\partial/\partial x_2$ and so we have

$$\begin{aligned}
&-\int_{x_2=\ell}\left[\psi_n\frac{\partial\Psi}{\partial x_2}-\Psi\frac{\partial\psi_n}{\partial x_2}\right]\mathrm{d}y_2\\
&\quad=\mathrm{e}^{\mathrm{i}\lambda(\ell-x_1)}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\beta y_{21}}\left[-\mathrm{i}\lambda\psi_n+\cos\alpha_{12}\frac{\partial\psi_n}{\partial R_{12}}\right.\\
&\qquad\left.-\frac{\sin\alpha_{12}}{R_{12}}\frac{\partial\psi_n}{\partial\alpha_{12}}\right]_{x_2=\ell}\mathrm{d}y_2\\
&\quad=\mathrm{e}^{\mathrm{i}\lambda(\ell-x_1)}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\beta y_{21}}\\
&\qquad\times\left[-\mathrm{i}\lambda\psi_n+\frac{k}{2}(\psi_{n-1}-\psi_{n+1})\right]_{x_2=\ell}\mathrm{d}y_2\\
&\quad=\frac{2}{\alpha}\mathrm{e}^{\mathrm{i}(\alpha-\lambda)(x_1-\ell)}\mathrm{i}^{n-1}\mathrm{e}^{\mathrm{i}n\theta_{\mathrm{in}}}(\lambda+\alpha),
\end{aligned} \tag{67}$$

using $2H_n'(x)=H_{n-1}(x)-H_{n+1}(x)$, $(2n/x)H_n(x)=H_{n-1}(x)+H_{n+1}(x)$, and Eq. (66) thrice.

The contribution from the circle $R_{12}=b$ is

$$\begin{aligned}
&-\int_0^{2\pi}\left[\psi_n\frac{\partial}{\partial R}(\mathrm{e}^{\mathrm{i}KR\cos(\Theta-\varphi)})-\mathrm{e}^{\mathrm{i}KR\cos(\Theta-\varphi)}\frac{\partial\psi_n}{\partial R}\right]_{R=b}b\mathrm{d}\Theta\\
&\quad=-b\int_0^{2\pi}\mathrm{e}^{\mathrm{i}Kb\cos(\Theta-\varphi)}\mathrm{e}^{\mathrm{i}n\Theta}\\
&\qquad\times[\mathrm{i}KH_n(kb)\cos(\Theta-\varphi)-kH_n'(kb)]\mathrm{d}\Theta\\
&\quad=-b\mathrm{e}^{\mathrm{i}n\varphi}\int_0^{2\pi}\mathrm{e}^{\mathrm{i}n\theta}\sum_{q=-\infty}^{\infty}\mathrm{i}^qJ_q(Kb)\mathrm{e}^{-\mathrm{i}q\theta}\\
&\qquad\times\left[\frac{\mathrm{i}K}{2}H_n(kb)(\mathrm{e}^{\mathrm{i}\theta}+\mathrm{e}^{-\mathrm{i}\theta})-kH_n'(kb)\right]\mathrm{d}\theta\\
&\quad=-2\pi b\mathrm{i}^n\mathrm{e}^{\mathrm{i}n\varphi}[KH_n(kb)J_n'(Kb)-kH_n'(kb)J_n(Kb)].
\end{aligned} \tag{68}$$

Thus, the system (59) can be written as

$$\begin{aligned}
&\mathcal{A}_m\mathrm{e}^{-\mathrm{i}m\varphi}\mathrm{e}^{\mathrm{i}\lambda x}+\mathcal{B}\mathrm{e}^{-\mathrm{i}m\theta_{\mathrm{in}}}\mathrm{e}^{\mathrm{i}\alpha x}\\
&\quad=-\mathrm{e}^{-\mathrm{i}m\theta_{\mathrm{in}}}\mathrm{e}^{\mathrm{i}\alpha x},\quad x>\ell+b,\quad m\in\mathbb{Z},
\end{aligned} \tag{69}$$

where

$$\mathcal{A}_m=F_m+\frac{2n_0\pi}{k^2-K^2}\sum_{n=-\infty}^{\infty}F_nZ_n\mathcal{N}_{n-m}(Kb),$$

$$\mathcal{B}=\frac{2n_0}{\alpha}\sum_{n=-\infty}^{\infty}Z_n\mathrm{e}^{\mathrm{i}n\theta_{\mathrm{in}}}\times\left\{\int_0^{\ell}\Phi_n(t)\mathrm{e}^{-\mathrm{i}\alpha t}\mathrm{d}t+\frac{\mathrm{i}F_n\mathrm{e}^{-\mathrm{i}n\varphi}}{\lambda-\alpha}\mathrm{e}^{\mathrm{i}(\lambda-\alpha)\ell}\right\},$$

and

$$\mathcal{N}_n(Kb)=kbH_n'(kb)J_n(Kb)-KbH_n(kb)J_n'(Kb). \tag{70}$$

In particular, note that $\mathcal{N}_0$ appeared in Sec. III C during our analysis of Lax's integral equation.

From Eq. (69), we immediately obtain $\mathcal{B}=-1$ and $\mathcal{A}_m=0$ for all m; the second of these, namely

$$F_m+\frac{2n_0\pi}{k^2-K^2}\sum_{n=-\infty}^{\infty}F_nZ_n\mathcal{N}_{n-m}(Kb)=0,\quad m\in\mathbb{Z}, \tag{71}$$

is of most interest to us. It is an infinite homogeneous system of linear algebraic equations for F_m, $m\in\mathbb{Z}$. The existence of a nontrivial solution to Eq. (71) determines K. Notice that Eq. (71) does not depend on θ_{in}, so that the effective wavenumber cannot depend on θ_{in}.

Equation (71) is the same as Eq. (33) in Bose and Mal (1973) [with the choice Eq. (14)]; these authors began by considering normal incidence, $\theta_{\mathrm{in}}=0$. However, the *derivation* of Eq. (71) given here has some advantages over that given by Bose and Mal (1973). First, we do not invoke "the so-called 'extinction theorem'" of Lax; this is described in Sec. VI of Lax (1952). Roughly speaking, this "theorem" asserts that one may simply delete the incident field when calculating the effective wavenumber, in the limit $N\to\infty$. Along with this come some divergent integrals; for example, the integrals in the unnumbered equation between Eqs. (32)

and (33) of Bose and Mal (1973) are divergent, because e^{iKx} is exponentially large as $x\to-\infty$. In fact, we can say that our analysis *proves* Lax's theorem in our particular case.

Second, when dealing with a half-space containing scatterers, we know from the work of Lloyd and Berry (1967) that the boundary of the half-space can cause difficulties. Here, we give a proper treatment of this boundary. In particular, we do not assume that all fields are proportional to $e^{i\lambda x}$ *everywhere* inside the half-space, $x>0$, but only in $x>\ell$, away from the boundary: the width of the boundary layer, ℓ, is not specified, and need not be specified if one only wants to calculate K.

A more recent analysis was given by Siqueira and Sarabandi (1996). They allow noncircular and nonidentical cylinders (using a T-matrix formulation) but they do assume that the effective field is proportional to $e^{i\lambda x}$ for all $x>0$.

C. Approximate determination of K for small n_0

The only approximation made in the derivation of Eq. (71) is the QCA, which is expected to be valid for small values of the scatterer concentration ($n_0a^2\ll1$). We now assume (as in Sec. III C) that n_0/k^2 is also small and write $K^2=k^2+\delta_1n_0+\delta_2n_0^2+\cdots$. We then have

$$\mathcal{N}_n(Kb)=2i/\pi+\tfrac{1}{2}b^2\delta_1d_n(kb)n_0+O(n_0^2), \tag{72}$$

where

$$d_n(x)=J_n'(x)H_n'(x)+[1-(n/x)^2]J_n(x)H_n(x) \tag{73}$$

and so

$$\frac{\mathcal{N}_n(Kb)}{k^2-K^2}=-\frac{2i}{\pi\delta_1n_0}-\frac{b^2d_n(kb)}{2}+\frac{2i\delta_2}{\pi\delta_1^2}+O(n_0). \tag{74}$$

If Eq. (74) is substituted in Eq. (71) and $O(n_0^2)$ terms neglected we get

$$F_m=\frac{4i}{\delta_1}\sum_{n=-\infty}^{\infty}Z_nF_n+n_0\sum_{n=-\infty}^{\infty}Z_nF_n\times\left(\pi b^2d_{n-m}(kb)-\frac{4i\delta_2}{\delta_1^2}\right),\quad m\in\mathbb{Z}. \tag{75}$$

At leading order this gives

$$F_m=\frac{4i}{\delta_1}\sum_{n=-\infty}^{\infty}Z_nF_n,\quad m\in\mathbb{Z}, \tag{76}$$

which implies that all the F_m are equal. If we write $F_m=F$, Eq. (76) becomes

$$\delta_1=4i\sum_{s=-\infty}^{\infty}Z_s=-4if(0), \tag{77}$$

where f is the far-field pattern, given by Eq. (53).

Returning to Eq. (75), we now put $F_m=F+n_0q_m$, and then the $O(n_0)$ terms give

$$q_m=-\frac{1}{f(0)}\sum_{n=-\infty}^{\infty}Z_nq_n+\pi b^2F\sum_{n=-\infty}^{\infty}Z_nd_{n-m}(kb)-\frac{iF\delta_2}{4f(0)},\quad m\in\mathbb{Z}. \tag{78}$$

It follows that $q_m-\pi b^2F\Sigma_{n=-\infty}^{\infty}Z_nd_{n-m}$ must be independent of m, call it Q:

$$Q=-\frac{1}{f(0)}\sum_{n=-\infty}^{\infty}Z_nq_n-\frac{iF\delta_2}{4f(0)}=-\frac{1}{f(0)}\sum_{n=-\infty}^{\infty}Z_n\left(Q+F\pi b^2\sum_{s=-\infty}^{\infty}Z_sd_{s-n}(kb)\right)-\frac{iF\delta_2}{4f(0)}. \tag{79}$$

Hence

$$\delta_2=4\pi ib^2\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}Z_nZ_sd_{s-n}(kb) \tag{80}$$

and so we obtain the approximation

$$K^2=k^2-4in_0f(0)+4\pi ib^2n_0^2\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}Z_nZ_sd_{s-n}(kb)+\cdots. \tag{81}$$

For isotropic point scatterers, we have $|Z_0|\gg|Z_n|$ for all $n\neq0$ and $g=-Z_0$, so that Eq. (81) reduces to Eq. (42) in this limit.

So far we have not made any assumptions about the size of ka or kb (though clearly $kb\geqslant2ka$). Now we will assume that kb is small. In the limit $x\to0$, we have $x^2d_n(x)\sim2i|n|/\pi$. Hence as $kb\to0$,

$$\delta_2\sim-\frac{8}{k^2}\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}|s-n|Z_nZ_s. \tag{82}$$

Now

$$[f(\theta)]^2=\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}Z_nZ_se^{i(n+s)\theta}=\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}Z_nZ_s\cos(n-s)\theta \tag{83}$$

since $Z_n=Z_{-n}$. Thus

$$\frac{d}{d\theta}[f(\theta)]^2=-\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}(n-s)Z_nZ_s\sin(n-s)\theta. \tag{84}$$

Also

$$\int_0^{\pi}\cot\tfrac{1}{2}\theta\sin m\theta\,d\theta=\pi\,\mathrm{sgn}(m), \tag{85}$$

see Eq. 3.612(7) in Gradshteyn and Ryzhik (2000). Thus, setting $kb=0$ gives

$$K^2=k^2-4in_0f(0)+\frac{8n_0^2}{\pi k^2}\int_0^{\pi}\cot(\theta/2)\frac{\mathrm{d}}{\mathrm{d}\theta}[f(\theta)]^2\mathrm{d}\theta. \tag{86}$$

The integral appearing here is convergent because $f'(0)=0$.

D. Effect of pair-correlation function choice

Here, we consider the effect of using a more complicated pair-correlation function, defined by Eq. (15) in terms of the function $\chi(r;n_0)$. This function must decay rapidly to zero as $r\to\infty$ and, in addition, $\chi(r;n_0)\to 0$ as $n_0\to 0$ for any fixed r. For example, Bose and Mal (1973) suggest using $\chi(r;n_0)=\mathrm{e}^{-r/L(n_0)}$, where the correlation length $L(n_0)\to 0$ as $n_0\to 0$. Other authors have supposed that $\chi(r;n_0)=0$ for $r>b'>b$, where the radius b' may be taken as $2b$; see, for example, p. 1072 of Bose (1996) or Eq. (27) in Twersky (1978).

Proceeding as in Sec. IV B, we obtain Eq. (57) with an additional factor of $[1+\chi(R_{12};n_0)]$ in the integrand. Evaluating this equation for $x_1>\ell+b'$, assuming that $\chi(r;n_0)=0$ for $r>b'$, we obtain Eq. (59) with M_{n-m} replaced by M'_{n-m}, where

$$\begin{aligned}M'_n&=M_n+\int_{b<R_{12}<b'}\psi_n(x_{21},y_{21})\\&\quad\times\Psi(x_{21},y_{21})\chi(R_{12};n_0)\mathrm{d}x_2\mathrm{d}y_2\\&=M_n+2\pi\mathrm{i}^n\mathrm{e}^{\mathrm{i}n\varphi}W_n,\\W_n&=\int_b^{b'}H_n(kR)J_n(KR)\chi(R;n_0)R\mathrm{d}R,\end{aligned}$$

and M_n is defined by Eq. (61). Hence, we obtain a modified form of Eq. (71), namely

$$F_m+2n_0\pi\sum_{n=-\infty}^{\infty}F_nZ_n\left\{\frac{\mathcal{N}_{n-m}(Kb)}{k^2-K^2}+W_{n-m}\right\}=0,\quad m\in\mathbb{Z}, \tag{87}$$

from which K is to be determined. This homogeneous system for F_n is Eq. (33) in Bose and Mal (1973) and it is a special case of Eq. (24) in Siqueira and Sarabandi (1996). Moreover, the fact that $W_n=o(1)$ as $n_0\to 0$ means that the approximations for K obtained in Sec. IV C, namely Eqs. (81) and (86), are unchanged by the presence of χ.

E. Reproducing Twersky's formula

It is implicit in the work of Twersky (1962) (and others) that the complications arising when a scatterer center is closer to the boundary $x=0$ than its radius are ignored. It was pointed out by Lloyd and Berry (1967) that, since all scatterers are treated equally, ignoring the boundary-layer effects is equivalent to using a pair-correlation function with the following property: if one scatterer is at (x_1,y_1), then no other scatterer [with center (x_2,y_2)] can occupy the infinite strip $x_1-a<x_2<x_1+a$. Thus, instead of Eq. (14), the choice

$$Np(\mathbf{r}_2|\mathbf{r}_1)=\begin{cases}0, & |x_{21}|<a,\\ n_0, & |x_{21}|>a,\end{cases} \tag{88}$$

was made. We shall show that use of Eq. (88) leads to Twersky's formula, Eq. (4).

Setting $s=1$ and taking the conditional average of Eq. (52) in the usual way, and looking for a solution in the form of Eq. (56) now leads to

$$\begin{aligned}&\Phi_m(x_1)+n_0\sum_{n=-\infty}^{\infty}Z_n(-\mathrm{i})^{n-m}\left(\int_0^{x_1-a}+\int_{x_1+a}^{\infty}\right)\\&\quad\times L_{n-m}(x_{21})\Phi_n(x_2)\mathrm{d}x_2=-\mathrm{e}^{-\mathrm{i}m\theta_{\mathrm{in}}}\mathrm{e}^{\mathrm{i}\alpha x_1},\quad m\in\mathbb{Z},\end{aligned} \tag{89}$$

where $L_n(X)$ is defined by Eq. (60).

Suppose that for $x>\ell$ we can write [cf. Eq. (58)]

$$\Phi_m(x)=F_m\mathrm{e}^{-\mathrm{i}m\theta_{\mathrm{in}}}\mathrm{e}^{\mathrm{i}\lambda x}, \tag{90}$$

where $\mathrm{Im}\,\lambda>0$. Then if $x_1>\ell+a$, Eq. (89) becomes

$$\begin{aligned}&F_m\mathrm{e}^{\mathrm{i}\lambda x_1}+n_0\sum_{n=-\infty}^{\infty}Z_n(-\mathrm{i})^{n-m}\mathrm{e}^{\mathrm{i}m\theta_{\mathrm{in}}}\int_0^{\ell}\Phi_n(x_2)L_{n-m}(x_{21})\\&\quad\times\mathrm{d}x_2+n_0\mathrm{e}^{\mathrm{i}\lambda x_1}\sum_{n=-\infty}^{\infty}F_nZ_n\mathrm{e}^{-\mathrm{i}(n-m)(\pi/2+\theta_{\mathrm{in}})}\\&\quad\times\left(\int_{\ell}^{x_1-a}+\int_{x_1+a}^{\infty}\right)L_{n-m}(x_{21})\mathrm{e}^{\mathrm{i}\lambda x_{21}}\mathrm{d}x_2=-\mathrm{e}^{\mathrm{i}\alpha x_1},\\&\quad m\in\mathbb{Z}.\end{aligned} \tag{91}$$

We have already evaluated $L_n(x)$ for $x<0$, see Eq. (66). Now, we also need its value for $x>0$; we have

$$\alpha L_n(x)=\begin{cases}2(-\mathrm{i})^n\mathrm{e}^{-\mathrm{i}n\theta_{\mathrm{in}}}\mathrm{e}^{\mathrm{i}\alpha x} & x>0\\ 2\mathrm{i}^n\mathrm{e}^{\mathrm{i}n\theta_{\mathrm{in}}}\mathrm{e}^{-\mathrm{i}\alpha x} & x<0.\end{cases} \tag{92}$$

Using these in Eq. (91) gives

$$\tilde{\mathcal{A}}_m\mathrm{e}^{\mathrm{i}\lambda x}+\tilde{\mathcal{B}}\mathrm{e}^{\mathrm{i}\alpha x}=-\mathrm{e}^{\mathrm{i}\alpha x},\quad x>\ell+a,\quad m\in\mathbb{Z},$$

where

$$\begin{aligned}\tilde{\mathcal{A}}_m&=F_m-\frac{2in_0}{\alpha}\sum_{n=-\infty}^{\infty}F_nZ_n\\&\quad\times\left\{\frac{\mathrm{e}^{-\mathrm{i}(\lambda-\alpha)a}}{\lambda-\alpha}-\frac{\mathrm{e}^{\mathrm{i}(\lambda+\alpha)a}}{\lambda+\alpha}\mathrm{e}^{\mathrm{i}(n-m)\theta_{\mathrm{T}}}\right\},\\\tilde{\mathcal{B}}&=\frac{2n_0}{\alpha}\sum_{n=-\infty}^{\infty}Z_n\left\{\mathrm{e}^{\mathrm{i}n\theta_{\mathrm{in}}}\int_0^{\ell}\Phi_n(t)\mathrm{e}^{-\mathrm{i}\alpha t}\mathrm{d}t+\frac{\mathrm{i}F_n\mathrm{e}^{\mathrm{i}(\lambda-\alpha)\ell}}{\lambda-\alpha}\right\}\end{aligned}$$

and $\theta_{\mathrm{T}}=\pi-2\theta_{\mathrm{in}}$. Thus, λ is to be found from $\tilde{\mathcal{A}}_m=0$ for all $m\in\mathbb{Z}$.

As before, we write $K^2-k^2=\lambda^2-\alpha^2=\delta_1n_0+\delta_2n_0^2+\cdots$. Hence,

$$\begin{aligned}&\frac{\mathrm{e}^{-\mathrm{i}(\lambda-\alpha)a}}{\lambda-\alpha}-\frac{\mathrm{e}^{\mathrm{i}(\lambda+\alpha)a}}{\lambda+\alpha}\mathrm{e}^{\mathrm{i}(n-m)\theta_{\mathrm{T}}}\\&\quad=\frac{2\alpha}{\delta_1n_0}+\frac{1}{2\alpha}\{1-2\mathrm{i}\alpha a-\mathrm{e}^{2\mathrm{i}\alpha a}\mathrm{e}^{\mathrm{i}(n-m)\theta_{\mathrm{T}}}\}-\frac{2\alpha\delta_2}{\delta_1^2}\\&\quad\quad+O(n_0).\end{aligned}$$

Substituting in $\tilde{\mathcal{A}}_m = 0$ and neglecting terms that are $O(n_0^2)$, we obtain

$$F_m - \sum_{n=-\infty}^{\infty} F_n Z_n \left\{ \frac{4\mathrm{i}}{\delta_1} + \frac{\mathrm{i}n_0}{\alpha^2}\{1 - 2\mathrm{i}\alpha a - \mathrm{e}^{2\mathrm{i}\alpha a}\mathrm{e}^{\mathrm{i}(n-m)\theta_\mathrm{T}}\} - \frac{4\mathrm{i}n_0\delta_2}{\delta_1^2} \right\} = 0$$

for $m \in \mathbb{Z}$. Proceeding as in Sec. IV C, we obtain Eq. (77), as before. Then, the $O(n_0)$ terms give

$$\delta_2 = \frac{\delta_1^2}{4\alpha^2}(1 - 2\mathrm{i}\alpha a) + \frac{4}{\alpha^2}\mathrm{e}^{2\mathrm{i}\alpha a}\sum_{n=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} Z_n Z_s \mathrm{e}^{\mathrm{i}(s-n)\theta_\mathrm{T}} = \frac{4}{\alpha^2}\{\mathrm{e}^{2\mathrm{i}\alpha a}[f(\theta_T)]^2 - (1 - 2\mathrm{i}\alpha a)[f(0)]^2\}.$$

Hence, if we let $\alpha a \to 0$ in this formula, we recover Twersky's (erroneous) formula, Eq. (4).

V. CONCLUDING REMARKS

We have derived a two-dimensional version of the three-dimensional Lloyd–Berry formula for the effective wave number in a dilute random configuration of scatterers, using methods that differ from those used by Lloyd and Berry (1967). Much remains to be done in order to validate the new formula. Specifically, it should be possible to compare its predictions with those obtained from full numerical simulations (using Monte Carlo methods) and from experiments. Some comparisons between Monte Carlo results and solutions of the infinite system (87), for various choices of χ, have been reported by Siqueira and Sarabandi (1996). They used lossy cylinders, and found good agreement for low area fractions, with little dependence on χ.

Price *et al.* (1988) have compared the predictions of Twersky's formula, Eq. (4), with experimental results obtained from sound propagation through three forests; they found "poor" agreement, but perhaps this could be attributed to errors in the formula and the crude approximation of an actual forest by a random array of sound-hard circular cylinders. For a recent review of the quantification of attenuation effects due to trees, see (Attenborough, 2002). Several other experimental studies, in the context of fiber-reinforced materials, are cited in the paper by Verbis *et al.* (2001).

In three dimensions, there is an extensive literature on comparisons between experiments, direct numerical simulations, and various theories, including the Lloyd–Berry formula; see, for example, Sec. 4.3.12 of Povey (1997), Hipp *et al.* (1999), and references therein. For low volume fractions and properly modelled spheres, the agreement is generally good: according to Povey (1997), p. 133, there is "a sizable body of evidence in support of the acoustic multiple scattering theory."

ACKNOWLEDGMENT

This work was supported by EPSRC Grant No. GR/S35585/01.

Aristégui, C., and Angel, Y. C. (**2002**). "New results for isotropic point scatterers: Foldy revisited," Wave Motion **36**, 383–399.

Attenborough, K. (**2002**). "Sound propagation close to the ground," Annu. Rev. Fluid Mech. **34**, 51–82.

Bose, S. K. (**1996**). "Ultrasonic plane SH wave reflection from a unidirectional fibrous composite slab," J. Sound Vib. **193**, 1069–1078.

Bose, S. K., and Mal, A. K. (**1973**). "Longitudinal shear waves in a fiber-reinforced composite," Int. J. Solids Struct. **9**, 1075–1085.

Bussink, P. G. J., Iske, P. L., Oortwijn, J., and Verbist, G. L. M. M. (**1995**). "Self-consistent analysis of elastic wave propagation in two-dimensional matrix-inclusion composites," J. Mech. Phys. Solids **43**, 1673–1690.

Dukhin, A. S., and Goetz, P. J. (**2001**). "Acoustic and electroacoustic spectroscopy for characterizing concentrated dispersions and emulsions," Adv. Colloid Interface Sci. **92**, 73–132.

Embleton, T. F. W. (**1966**). "Scattering by an array of cylinders as a function of surface impedance," J. Acoust. Soc. Am. **40**, 667–670.

Fikioris, J. G. (**1966**). "Multiple scattering from inhomogeneous media," IEEE Trans. Antennas Propag. **14**, 202–207.

Fikioris, J. G., and Waterman, P. C. (**1964**). "Multiple scattering of waves. II. 'Hole corrections' in the scalar case," J. Math. Phys. **5**, 1413–1420.

Foldy, L. L. (**1945**). "The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers," Phys. Rev. **67**, 107–119.

Gradshteyn, I. S., and Ryzhik, I. M. (**2000**). *Tables of Integrals, Series and Products*, 6th ed. (Academic, New York).

Groenenboom, J., and Snieder, R. (**1995**). "Attenuation, dispersion, and anisotropy by multiple scattering of transmitted waves through distributions of scatterers," J. Acoust. Soc. Am. **98**, 3482–3492.

Hipp, A. K., Storti, G., and Morbidelli, M. (**1999**). "On multiple-particle effects in the acoustic characterization of colloidal dispersions," J. Phys. D **32**, 568–576.

Ishimaru, A. (**1978**). *Wave Propagation and Scattering in Random Media* (Academic, New York), Vol. 2.

Kanaun, S. K. (**2000**). "Self-consistent methods in the problems of wave propagation through heterogeneous media," *Heterogeneous Media. Micromechanics Modeling Methods and Simulations*, edited by K. Markov and L. Preziosi (Birkhäuser, Boston), pp. 241–319.

Kanaun, S. K., and Levin, V. M. (**2003**). "Effective medium method in the problem of axial elastic shear wave propagation through fiber composites," Int. J. Solids Struct. **40**, 4859–4878.

Kim, J.-Y. (**2003**). "Antiplane shear wave propagation in fiber-reinforced composites," J. Acoust. Soc. Am. **113**, 2442–2445.

Kim, J.-Y., Ih, J.-G., and Lee, B.-H. (**1995**). "Dispersion of elastic waves in random particulate composites," J. Acoust. Soc. Am. **97**, 1380–1388.

Lax, M. (**1951**). "Multiple scattering of waves," Rev. Mod. Phys. **23**, 287–310.

Lax, M. (**1952**). "Multiple scattering of waves. II. The effective field in dense systems," Phys. Rev. **85**, 621–629.

Linton, C. M., and McIver, P. (**2001**). *Handbook of Mathematical Techniques for Wave/Structure Interactions* (Chapman & Hall/CRC, Boca Raton).

Lloyd, P., and Berry, M. V. (**1967**). "Wave propagation through an assembly of spheres IV. Relations between different multiple scattering theories," Proc. Phys. Soc. London **91**, 678–688.

Mal, A. K., and Knopoff, L. (**1967**). "Elastic wave velocities in two-component systems," J. Inst. Math. Appl. **3**, 376–387.

McClements, D. J., Fairley, P., and Povey, M. J. W. (**1990**). "Comparison of effective medium and multiple-scattering theories of predicting the ultrasonic properties of dispersions," J. Acoust. Soc. Am. **87**, 2244–2246.

Povey, M. J. W. (**1997**). *Ultrasonic Techniques for Fluids Characterization* (Academic, San Diego).

Price, M. A., Attenborough, K., and Heap, N. W. (**1988**). "Sound attenuation through trees: measurements and models," J. Acoust. Soc. Am. **84**, 1836–1844.

Siqueira, P. R., and Sarabandi, K. (**1996**). "Method of moments evaluation

of the two-dimensional quasi-crystalline approximation,'' IEEE Trans. Antennas Propag. **44**, 1067–1077.
Siqueira, P. R., Sarabandi, K., and Ulaby, F. T. (**1995**). ''Numerical simulation of scatterer positions in a very dense medium with an application to the two-dimensional Born approximation,'' Radio Sci. **30**, 1325–1339.
Spelt, P. D. M., Norato, M. A., Sangani, A. S., Greenwood, M. S., and Tavlarides, L. L. (**2001**). ''Attenuation of sound in concentrated suspensions: theory and experiments,'' J. Fluid Mech. **430**, 51–86.
Tsang, L., Kong, J. A., Ding, K.-H., and Ao, C. O. (**2001**). *Scattering of Electromagnetic Waves: Numerical Simulations* (Wiley, New York).
Twersky, V. (**1962**). ''On scattering of waves by random distributions. I. Free-space scatterer formalism,'' J. Math. Phys. **3**, 700–715.
Twersky, V. (**1978**). ''Acoustic bulk parameters in distributions of pair-correlated scatterers,'' J. Acoust. Soc. Am. **64**, 1710–1719.
Varadan, V. K., Ma, Y., and Varadan, V. V. (**1986**). ''Multiple scattering of compressional and shear waves by fiber-reinforced composite materials,'' J. Acoust. Soc. Am. **80**, 333–339.
Varadan, V. K., Varadan, V. V., and Pao, Y.-H. (**1978**). ''Multiple scattering of elastic waves by cylinders of arbitrary cross section. I. SH waves,'' J. Acoust. Soc. Am. **63**, 1310–1319.
Verbis, J. T., Kattis, S. E., Tsinopoulos, S. V., and Polyzos, D. (**2001**). ''Wave dispersion and attenuation in fiber composites,'' Computational Mech. **27**, 244–252.
Waterman, P. C., and Truell, R. (**1961**). ''Multiple scattering of waves,'' J. Math. Phys. **2**, 512–537.
Yang, R.-B., and Mal, A. K. (**1994**). ''Multiple scattering of elastic waves in a fiber-reinforced composite,'' J. Mech. Phys. Solids **42**, 1945–1968.

Longitudinal and lateral low frequency head wave analysis in soft media

Nicolás Benech[a] and Carlos A. Negreira[b]
Laboratorio de Acústica Ultrasonora, Instituto de Física, Facultad de Ciencias, Igua 4225, Montevideo, 11400, Uruguay

(Received 27 September 2004; revised 14 March 2005; accepted 29 March 2005)

This article studies the influence of the head wave in the lateral and longitudinal components of the displacements generated by the radiation of low-frequency elastic waves in an isotropic and homogeneous soft solid. Low-frequency shear waves are used to characterize elastic properties of soft tissues. In this context, it is useful to have a detailed study of the low-frequency wave field in this kind of material. A soft medium is characterized by the fact that the head wave is found in the source's axis. Even though its amplitude is small compared with the shear wave, it is possible to be observed experimentally by recording consecutive ultrasonic A-lines while the low-frequency wave propagates inside the medium. A standard one-dimensional speckle tracking technique is employed to measure the displacements. Experimental results were interpreted through the exact Green's function solution to the half-space problem. According to the theoretical and experimental analysis, the head wave and surface related terms in general contribute to the displacements in the low-frequency range. This article thoroughly analyzes and experimentally shows the contribution of the head wave for the lateral component, which is not fully addressed by the literature. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1920147]

PACS numbers: 43.20.Gp, 43.20.Hq, 43.20.Px [TDM] Pages: 3424–3431

I. INTRODUCTION

The use of low-frequency shear waves has proved to be a useful means to obtain quantitative information about the elastic characterization of soft biologic tissues, with the objective of detecting lesions.[1,2] The dynamic elastographic techniques based on ultrasound use a mechanic vibrator that excites the medium from the surface in a monochromatic[1–4] or transient[5] way.

In order to obtain quantitative information about the elasticity of the studied medium, the elastographic techniques are based on a constitutive model of it. In elastography, tissues are usually considered elastic media at a first approach.[6–8] The simplest model consists of assuming an isotropic, homogeneous, and perfectly elastic medium. Despite the simplification this involves, the works that have used this model yield encouraging results, showing locally homogeneous regions but with different values of Lamé's coefficients.[9] In order to go further in the development of the dynamic elastographic techniques, it is important to have a detailed study of the transient response of soft elastic media. A recent paper by Sandrin *et al.*[10] shows the importance of the coupling term on the elastographic techniques. The authors calculated the impulse response for the longitudinal component of the displacements along the axis of a circular source, through a simplified Green's function for a semi-infinite medium. The simplification consists of not taking into account any surface effect on the displacement field inside the medium. These effects include the Rayleigh surface wave and conversion modes of waves that propagate along the interface into bulk waves, as described by Lhémery.[11] Head waves are a characteristic result of this mode conversion, generated by a conversion in the compressional wave propagating along the interface into a bulk shear wave.[11,12]

Head waves have been studied previously, mainly in seismology[13] and in the ultrasonic field for nondestructive evaluations in conventional solid media.[12,14] Although there are exact solutions to the half-space problem, they are difficult to deal with analytically, so head waves have been analyzed mainly in the asymptotic limit of high frequencies.[12] These results cannot be applied to this research since it works with low frequencies and thus does not meet the conditions that these approximations require. Therefore, this article experimentally studies the head wave and its influence in the displacements generated by a line source, acting normally from the free surface of a tissue-mimicking phantom. The experiments are carried out using an ultrasonic transducer to measure displacements along the ultrasonic beam inside the phantom by a standard one-dimensional speckle tracking technique. The analysis is based on interpreting the experimental results with those foreseen by the exact Green's function (without any approximation) obtained through the Cagniard–de Hoop method. This allows for an identification of the contribution of the head wave and other superficial terms to the displacements. Separate studies for the longitudinal and lateral components are carried out so as to identify the contribution to each. The existing literature does not study the lateral component thoroughly, though it degrades the quality of the results of the longitudinal one.[15,16]

II. THEORY

Let us consider the semispace $z \geqslant 0$ as elastic, homogeneous, and isotropic, with Lamé's coefficients λ and μ, and

[a]Electronic mail: nbenech@fisica.edu.uy
[b]Electronic mail: carlosn@fisica.edu.uy

 0001-4966/2005/117(6)/3424/8/$22.50

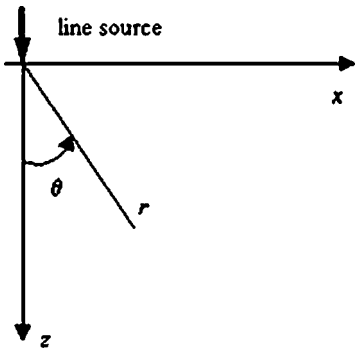


FIG. 1. Lamb's problem coordinates definition.

material density ρ. An infinite line source that generates the propagation of elastic waves in the medium acts from the free surface. Let us suppose that the line source coincides with the y axis and that it behaves impulsively in the z-axis direction (Fig. 1). The initial conditions are of null displacement and velocity for $t<0$. As usual, the problem is solved by using the vector potential $\boldsymbol{\psi}$ and the scalar potential φ, which are related to the displacement $\mathbf{u}$ by[13]

$$\mathbf{u}(\mathbf{r},t)=\nabla\varphi(\mathbf{r},t)+\nabla\times\boldsymbol{\psi}(\mathbf{r},t). \tag{1}$$

Due to symmetry considerations, the displacement $\mathbf{u}(\mathbf{r},t)$ neither depends on the y coordinate nor does it have a component in that direction. Hence, Eq. (1) can be written as

$$\mathbf{u}(x,z,t)=\nabla\varphi(x,z,t)+\nabla\times\psi(x,z,t)\mathbf{j}, \tag{2}$$

where $\boldsymbol{\psi}$ is substituted by: $\boldsymbol{\psi}=\psi\mathbf{j}$, with $\mathbf{j}$ being a unit vector in the positive direction of the y axis. Therefore, the problem is bidimensional and the displacement's study is made along the $x-z$ plane. This problem can be solved in an exact way using the Cagniard–de Hoop method.[13] Results are shown in the Appendix. The final solution, which is noted as $G_\alpha(x,z,t)$, where $\alpha=x$, z indicates the displacement component, can be written as the addition of three terms

$$G_\alpha(x,z,t)=G_\alpha^P(x,z,t)+G_\alpha^S(x,z,t)+G_\alpha^{HW}(x,z,t). \tag{3}$$

The first of these terms is associated with the compressional wave or P wave, the second with the shear wave or S wave, and the third with the head wave. This wave, which originates in the alteration of the integration path when reversing the Laplace transform due to branch points, assures that the border conditions are satisfied, and is interpreted as a change in mode of the surface P wave that enters the volume as an S wave.[11,12,17] The existence of a head wave is registered beginning at a critical angle, which is defined by $\sin(\theta_c)=\gamma^{-1}$, where $\gamma=c_p/c_s$ is the quotient between the speeds of the bulk P- and S waves, respectively. For soft solids $c_p \gg c_s$, and hence the critical angle can be considered as being zero. As a consequence, the head wave exists in the whole medium. The arrival time of the head wave depends on the angle θ with respect to the z axis

$$t_{hw}=\frac{r}{c_p}[(\gamma^2-1)^{1/2}\cos(\theta)+\sin(\theta)]$$
$$\cong t_p[\gamma\cos(\theta)+\sin(\theta)]. \tag{4}$$

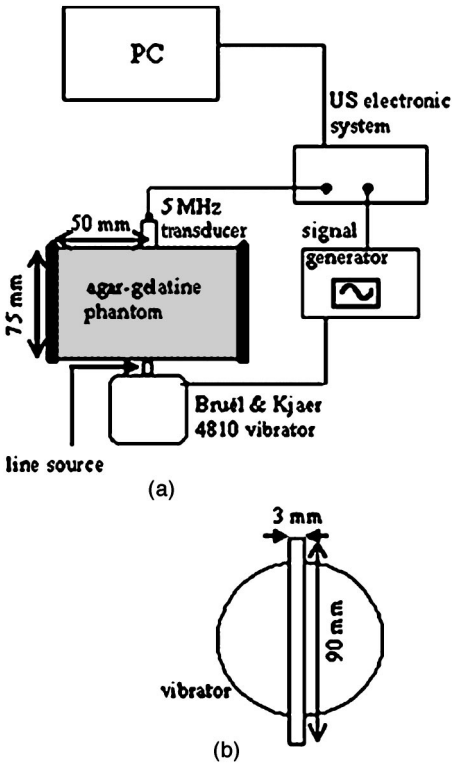


FIG. 2. (a) Experimental setup. (b) Dimensions of the line source employed.

In the last expression, the substitution r/c_p by t_p, the arrival time of the compressional wave, was made, and the fact that $\gamma\gg1$ was used to simplify the square root. Depending on the angular position, t_{hw} varies between t_s for $\theta=\theta_c$ and t_p for $\theta=\pi/2$. In $t=t_{hw}$, the solution does not show a discontinuity but a soft behavior. However, this term keeps on going until $t_s=r/c_s$, the arrival time of the shear wave, where, as the shear wave, it features a singularity. These singularities are integrable, so that there is no difficulty in obtaining the response due to an arbitrary time function source by convolution. It is difficult to deal with the analytic expression of the head wave, so numerical calculations were performed to detect its contribution to displacements, showing it is small compared to the shear wave, but even then, detectable with our experimental setup.

III. EXPERIMENTAL SETUP

Figure 2 shows the experimental setup. The material used for this experiment is a tissue-mimicking phantom, made of agar-gelatin. The phantoms built in this way behave like elastic media in a first approach. The attenuation measured for the shear waves is of the order of 15 m^{-1} for frequencies between 60 and 120 Hz. This phantom, prepared in a cylinder-like container 100 mm in diameter and 75 mm in height, is put between a 5-MHz central frequency ultrasonic transducer 10 mm in diameter, and a Bruel & Kjaer 4810 vibrator to which is attached a 9.0-mm-long and 3-mm-wide line [Fig. 2(b)] that vibrates normally with respect to the phantom's free surface. The length was chosen so that the line would behave as infinite near its center, and the width so that it would be much smaller than the distance to the source at which measurements are made (30 mm) and thus be con-

sidered a line source. Line sources are employed in elastography to focus shear waves in the symmetry plane between them.[9,18] Therefore, the wave-field analysis generated by a single line is a contribution to this kind of application. Using a function generator, a signal is send to the vibrator synchronized with the ultrasonic pulses in the emission–reception mode. As excitation, we use two sinusoidal cycles of central frequency f_0, which is changed in the experiments. The ultrasonic transducer is employed to record several speckle signals from the medium while the low-frequency wave is propagating on it. The acquisition system allows for a registration up to 512 consecutive ecographic lines of 2048 points each, sampled at 40 MHz, with a repetition frequency of 1 KHz. This way, each signal contains information about the position of the scatterers inside the medium at regular times of 1 ms. Displacements are measured along the ultrasonic beam using a standard correlation technique among the consecutive ultrasonic A lines.[19] This technique allows for measurement of the displacements that take place between the time interval of two consecutive ultrasonic pulses, so what we actually measure is the particle velocity.[20] The accuracy of the measurements depends on the transducer frequency, the signal-to-noise ratio, and the time width of the windows that were selected in the correlation algorithm. In this experience, we use a 5-MHz central frequency transducer; the width of the windows is 2.4 μs (6 wavelengths) with a 50% overlap among them. The calculated standard deviation using the Cramér–Rao lower bound, obtained by Walker and Trahey,[21] for these experimental conditions is of the order of 1 μm.

The displacements due to an arbitrary excitation of time $f(t)$ are obtained as the convolution product between $f(t)$ and the Green's function. Since what is measured experimentally is actually the particle velocity, the time derivative of the Green's function is used

$$v_\alpha(x,z,t) = \frac{\partial G_\alpha(x,z,t)}{\partial t} \otimes f(t). \tag{5}$$

IV. RESULTS

A. Longitudinal component

The longitudinal component is the most relevant in ultrasonic elastography because it is the component which we access directly in the experiments so it forms the basis to deduce the elastic properties of the medium. Although the head wave is present in the whole medium, we cannot separate it from the shear wave for small angles. In view of Eq. (4), t_{hw} tends to t_s as θ tends to θ_c. In order to interpret experimental results through the Green's function analysis, we plot in Fig. 3(a) its time dependence in the z axis 30 mm from the source using experimental values for the speed of *P*- and *S* waves, 1500 and 2.8 m/s, respectively. Due to the directivity pattern of each term,[22] the only relevant contribution in the z axis is that associated with the G_z^P term. Even though it is not explicit in the exact Green's function, this term contains the coupling term. It shows a Dirac delta in $t = t_p$ of very small amplitude, corresponding to the arrival of the compressional wave, and then a ramp that grows until $t = t_s$, the coupling term. After the end of the ramp there is a tail attributed to surface phenomena as noted by Sandrin *et al.*[10] The slope of the ramp is not constant (as expected from the coupling term alone) but it features a slow variation between t_p and t_s as shown in Fig. 3(b). This slope variation also is due to surface phenomena. These two effects are not associated with the head wave (strictly the head wave is null on axis) but to other wave-field coupling phenomena at the free surface. These terms, together with the coupling term, appear in the Green's function spectrum in the low-frequency range. Figure 4 shows the module of the Fourier transform of the time derivative of the Green's function shown in Fig. 3(a). Present on it are several maxima and

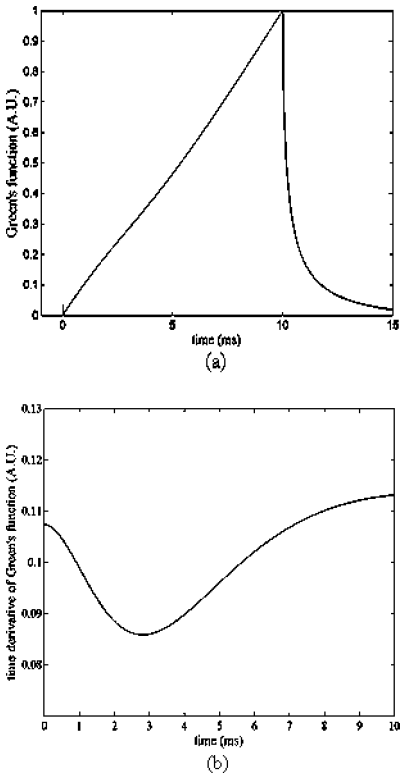


FIG. 3. Time dependence of Green's function (a) and time derivative of Green's function (b) for a point on the z axis at 30 mm from the source. Time in (b) corresponds to values between t_p and t_s.

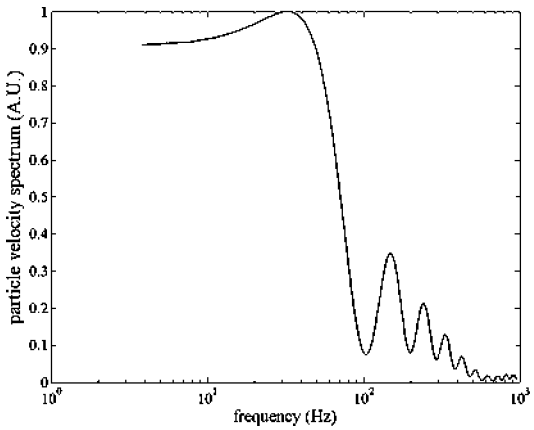


FIG. 4. Fourier transform modulus for the time derivative of Green's function. The low-frequency contribution comes from the coupling and superficial terms.

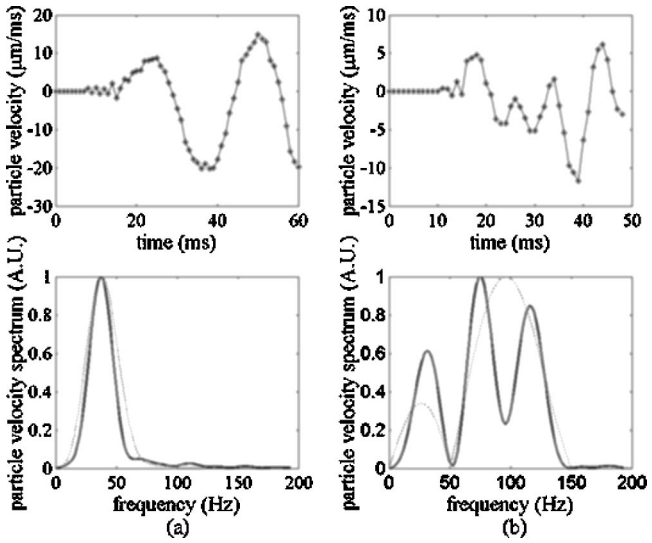


FIG. 5. Particle velocity (up) and corresponding spectra (below) measured on the *z* axis at 30 mm from the source for 40-Hz (a) and 100-Hz (b) central frequency excitations. The spectra at 30 mm (continuous) are compared with those corresponding to the excitation functions recorded in the surface (dotted). For the 40-Hz excitation it remains practically unchanged, but for the 100-Hz one the central frequency is modified due to interference between the *P* and *S* waves.

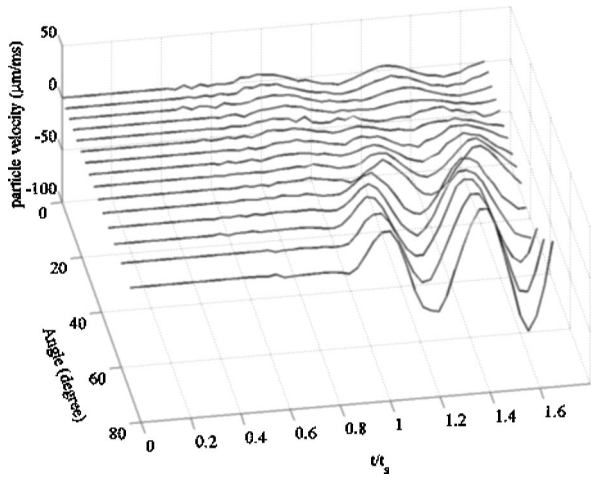


FIG. 6. Longitudinal particle velocity in function of normalized time t/t_s for several angles. As the angle increases, the *S* wave takes more relevance by virtue of its directivity pattern.

minima due to constructive and destructive interference between the compression and shear waves.[20] The time difference between these waves is approximately t_s, which for this case ($r=30$ mm and $c_s=2.8$ m/s) is $t_s=10.7$ ms. So, the first maxima is around 50 Hz and the first minima at 100 Hz. Below 50 Hz the spectrum continues on in high values, due to both the coupling term and surface-related terms. Figure 5 shows the particle velocity measured on *z* axis at 30 mm from the source, for excitations of $f_0=40$- and 100-Hz central frequency and their corresponding spectra. The spectra of the excitation functions recorded at the surface are also plotted in order to show their evolution as the wave propagates through the medium. These results agree with the previous analysis. As expected from Fig. 4, the maximum displacement recorded for the 40-Hz excitation (20.3 μm) is bigger than that of the higher frequency (11.8 μm), and its spectrum is practically the same as that recorded at the surface with a single peak at 37 Hz. The spectrum corresponding to the 100-Hz excitation function in the surface has two peaks at 23- and 97 Hz, while at 30 mm from the source it has three peaks at 27, 70, and 117 Hz, going through a minimum at around 100 Hz because of the spectral form of the Green's function.

The previous analysis suggests that the existence of superficial terms can be identified through the particle velocity spectrum. In Fourier's domain, Eq. (5) is written as

$$\tilde{v}(x,z,\omega)=i\omega\tilde{G}_{\alpha}(x,z,\omega)\tilde{f}(\omega), \tag{6}$$

where *i* is the imaginary unit. Therefore, the particle velocity spectrum depends on the spectrum of the Green's function, which in turn depends on the angle θ, due to the directivity pattern of each term. In this article, the interest is centered in the analysis of the displacements that take place outside the *z* axis because the contribution of the head wave is not experimentally identified on it as seen above. As the angle increases, the amplitude of the shear wave term G_z^S becomes more relevant. Figure 6 shows the particle velocity as a function of normalized time t/t_s at a distance of 30 mm from the source for various angular positions. In the figure it is observed that the amplitude of the *S* wave, which has the opposite sign of the coupling term, increases with angle. In the zone closest to the *z* axis, the displacements in $t=t_s$ are negative, while for higher angles they are positive. In this latest zone, we get closer to the free surface and the amplitude increases due to the Rayleigh wave penetrating the medium.[20] The speed of this wave, c_R, is given by the Viktorov's formula[23]

$$\frac{c_R}{c_s}\cong\frac{0.718-(c_s/c_p)^2}{0.75-(c_s/c_p)^2}. \tag{7}$$

In our case, the relationship $c_s/c_p\to 0$ and therefore $c_R\cong 0.95c_s$. As a consequence of this nearness of velocities, it is impossible to separate these two waves with our acquisition system.

In addition, as the angle increases, so does the difference between t_{hw} and t_s, and the time duration of the head wave. Its contribution is small, but we can get a good record of it in the zone where the amplitude of the coupling term is negligible. Figure 7(a) shows the displacement as a function of time for a distance of 30 mm from the source and an angle of 55° for a 80-Hz central frequency excitation. For this angular position the coupling term becomes negligible and the contribution to the displacements for times preceding t_s comes from the head wave. The speed measured for the shear wave is 2.81 ± 0.06 m/s. The speed of the compressional wave is of the order of 1500 m/s. Therefore, the arrival time of the head wave, Eq. (4), is $t_{hw}\cong 6$ ms. The excitation begins at 5 ms; hence, the arrival of the head wave happens at 11 ms and the one of the shear wave at 16 ms. Arrival times of each wave are indicated with arrows in the figure, and we also compare the experimental result with the one obtained after convoluting the exact Green's function with the excitation imposed to the medium, which was experimentally recorded on the surface. Figure 7(b) shows a comparison between the spectra of

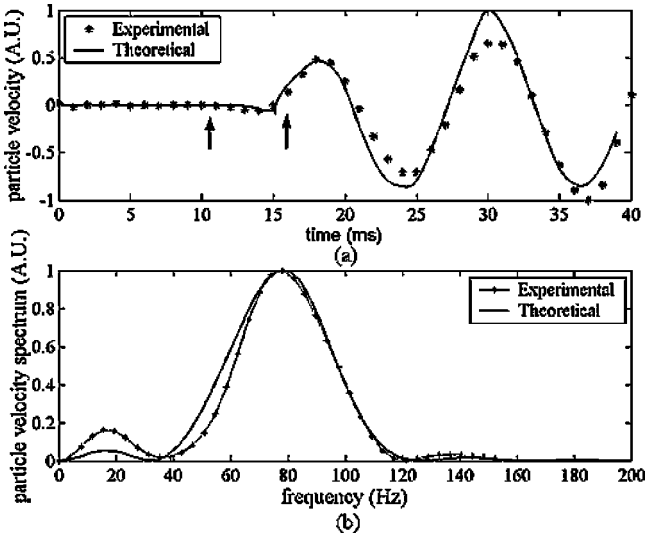


FIG. 7. (a) Experimental (asterisks) and theoretical (continuous) particle velocity as a function of time, where the contribution of the head wave is registered between the arrows. (b) Experimental particle velocity spectrum (asterisks) and theoretical (continuous) obtained through Eq. (6). Results are in good agreement with peaks at 20 and 74 Hz.

the experimental result with that obtained with Eq. (6), which agrees with the experimental result, with peaks at 20 and 74 Hz. Due to the time duration of the head wave (between t_{hw} and t_s), it always interacts with the *S* wave in making the convolution product. This interaction is identified as lowering the central frequency of the particle velocity spectrum. Figure 8 shows the frequencies of peaks in the Fourier transform modulus in function of angle for a 100-Hz central frequency excitation. These experimental results are compared with those obtained through the Green's function. We can observe a low-frequency peak for all angles in both results, showing that there are low-frequency terms in the Green's function spectrum for all angles. In addition, for small angles the spectrum shows the existence of three peaks, as shown in Fig. 5(b), due to interference phenomena. For greater angles, this interference is not present as a consequence of directivity pattern for each term as shown in Fig. 6, and the central frequency tends to be smaller than 100 Hz as the angle increases. In this zone, the coupling term is negligible and the head wave contributes to lower the central frequency of the spectrum.

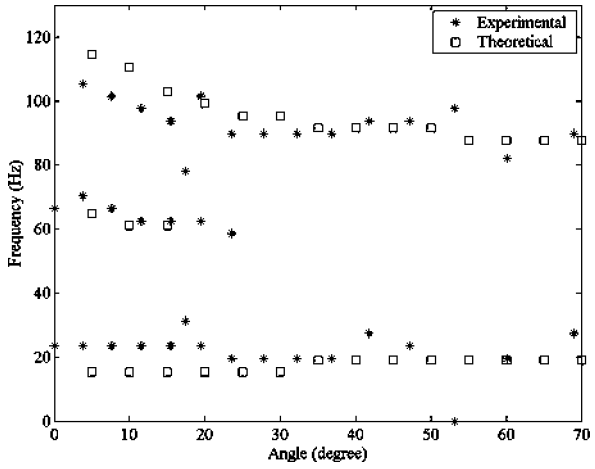


FIG. 8. Frequencies as a function of angle, corresponding to the peaks of the Fourier transform modulus for a 100-Hz central frequency excitation at 30 mm from the source. The experimental results (asterisks) are compared with those obtained through exact Green's function (squares). The central frequency decays as angle increases due to the head wave.

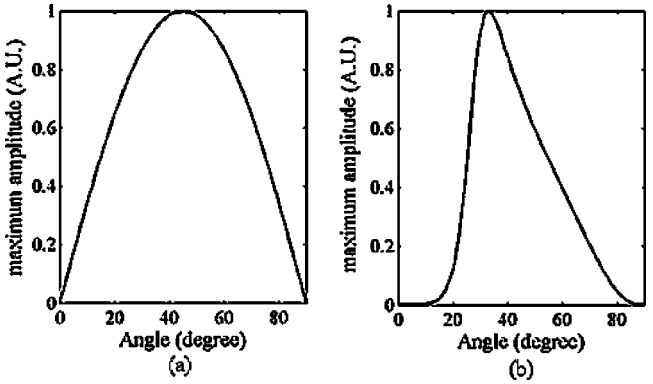


FIG. 9. Directivity pattern for the coupling term (a) and the *S* wave (b). The coupling term has its maximum at 45° and the *S* wave at 33°.

B. Lateral component

In the same way as what was exposed for the longitudinal component, the head wave contribution to the lateral component (*x* direction) of displacements is analyzed. Figure 9 shows the directivity patterns for the coupling term [Fig. 9(a)] and for the shear wave [Fig. 9(b)]. The coupling term has its maximum at 45° while for the shear wave it is at 33°, and displacements along the *z* axis are null for both terms. In these circumstances, the head wave and the coupling term are easy to confuse, and so head wave observation is delicate for the lateral component. Figure 10(a) shows the particle velocity as function of normalized time t/t_s, for various angular positions, where it is observed that the *S* wave dominates over the coupling term for all angles; Fig. 10(b) shows the maximum amplitude values of Fig. 10(a) as function of angle, reproducing the directivity pattern of *S* wave shown in Fig. 9(b). Directivity patterns were obtained experimentally before for the longitudinal component,[22] but not for the lateral one.

For small angles (less than 30°), the time separation between the head wave and the shear wave is less than 1 ms at 30 mm from the source, so it cannot be recorded clearly with our acquisition system. The contribution of the head wave between t_{hw} and t_s can be observed experimentally for greater angles, since in this zone the time separation with the shear wave increases and the amplitude of the coupling term decreases. Figure 11 shows the comparison between the experimental and theoretical lateral displacement at 30 mm from the source and an angle of 60°. In order to clearly identify the head wave contribution, and avoid confusion with the coupling term, we make here also a comparison of experimental results with those obtained with the simplified Green's function introduced by Sandrin *et al.*, adapted to a line source. As mentioned above, this function does not take into account any surface effect, so the head wave is not present in the particle velocity field. In Fig. 11(a) the comparison was made using the exact Green's function, while

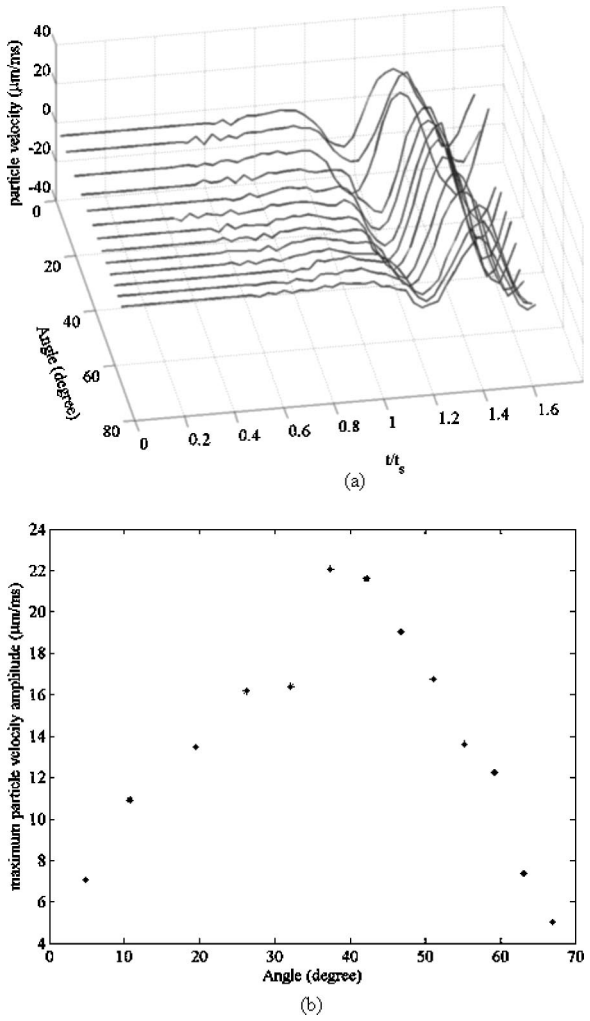


FIG. 10. (a) Lateral particle velocity in function of time for several angles. The *S* wave dominates for all angles. (b) Maximum amplitude in function of angle that experimentally confirms the directivity pattern of Fig. 9(b).

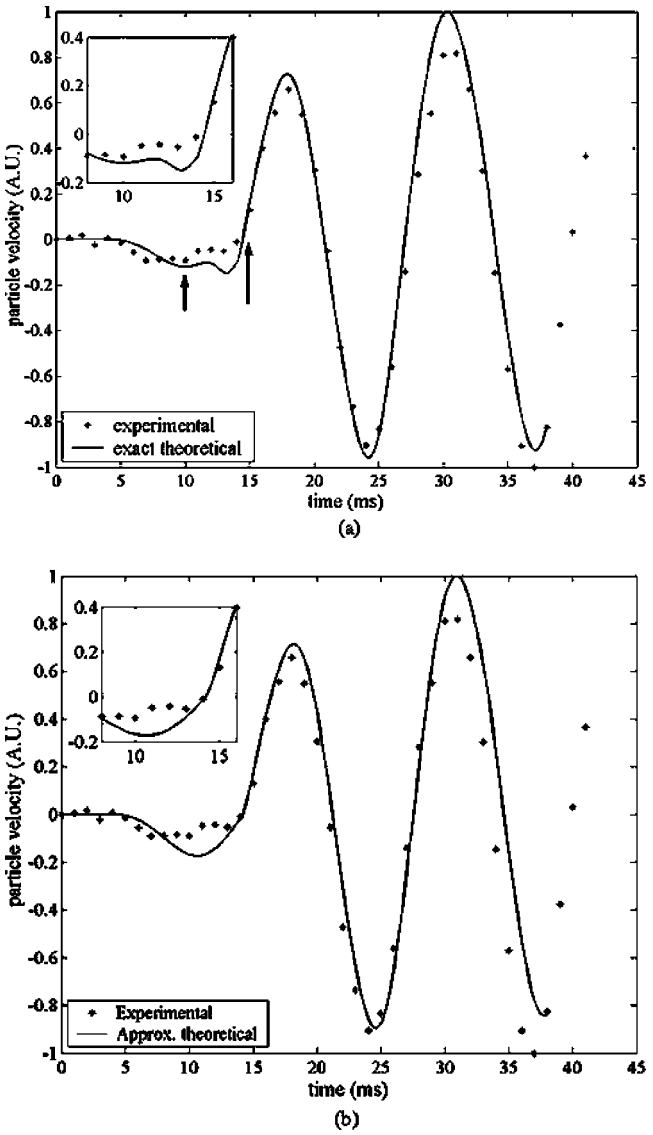


FIG. 11. Head wave contribution to the lateral particle velocity at 30 mm from the source and 60°. Experimental (asterisks) and theoretical (continuous) results obtained with the exact (a) and simplified (b) Green's function are compared. At this point, time separation between the head and the *S* waves is approximately 5 ms. The arrival time for each wave is indicated with arrows in (a). A zoom plot was added in each panel better show the zone previous to the shear wave arrival. Between these points the head wave features an oscillation in agreement with the experimental data, while it is not present in the continuous plot in (b).

Fig. 11(b) uses the simplified one. In both cases we observe the presence of the coupling term starting from t_p, but in Fig. 11(a) there is an oscillation between times t_{hw} and t_s (signaled by arrows in the figure) that shows an increase of amplitude after t_{hw} and then a decrease until the arrival of the shear wave, in agreement with the experimental data. Results that were obtained using the simplified Green's function do not have these oscillations, but the coupling term continues acting until the shear wave arrives. This figure shows that the head wave effect in the lateral particle velocity is small but detectable with our acquisition system.

V. CONCLUSIONS

In this article a detailed study of the low-frequency displacement field in soft media for the longitudinal and lateral component was made. Particularly, the experimental head wave contribution to the field is in good agreement with the theory of wave propagation in elastic media. Even though the head wave amplitude is small compared with the shear wave, the ultrasound-based experimental setup utilized was able to detect it and get a good record for the longitudinal and lateral component. The exact Green's function solution to the half-space problem was introduced to suitably take into account surface-related phenomena. Its analysis has yielded that surface terms are, in general, low-frequency range terms. This fact was reaffirmed by experimental results showing the presence of low-frequency terms for all angular positions and a decay in the central frequency as angle increases.

The experimental observation of the longitudinal component of the head wave is an important contribution to the development of elastography because the head wave becomes more relevant if extended sources are employed.[14] Another topic undertaken in this work was the lateral wavefield analysis, since it is not sufficiently studied in the literature. The radiation pattern obtained for this component is a

contribution, especially when using linear sources to focus shear waves in the symmetry plane between them.[18,24] We hereby propose future research on the influence of the head wave in elastography, specifically in the estimation of the velocity of the shear wave if we try to measure it outside the source's symmetry axis, and the spreading of this work to more realistic situations that could be found in some tissues such as anisotropic and viscoelastic media.

ACKNOWLEDGMENTS

We thank A. Kun and J. C. Benech, investigators of the Biophysics Department of the Institute of Biological Investigations Clemente Estable (IIBCE), for their participation in the tissue-mimicking phantom elaboration. This work was supported by the Programa para el Desarrollo de las Ciencias Básicas-Uruguay (PEDECIBA).

APPENDIX: EXACT GREEN'S FUNCTION

The exact solution to the Lamb's problem shown in Fig. 1 can be obtained by applying the Cagniard–de Hoop method. The final result is directly the impulse response, and is expressed as follows:

$$G_\alpha = G_\alpha^P + G_\alpha^S + G_\alpha^{HW}, \quad \alpha = x,z, \tag{A1}$$

where

$$G_\alpha^P(x,z,t) = \frac{H(t-t_p)}{\pi\mu}\,\mathrm{Re}\left[K_\alpha^P(q_p)\,\frac{dq_p}{dt}\right], \tag{A2}$$

$$G_\alpha^S(x,z,t) = \frac{H(t-t_s)}{\pi\mu}\,\mathrm{Re}\left[K_\alpha^S(q_s)\,\frac{dq_s}{dt}\right], \tag{A3}$$

$$G_\alpha^{HW}(x,z,t) = \frac{H(t-t_{hw})H(t_s-t)H(x/r-c_s/c_p)}{\pi\mu} \times \mathrm{Re}\left[K_\alpha^S(q_{hw})\,\frac{dq_{hw}}{dt}\right], \tag{A4}$$

where $H(t)$ is the Heaviside step function and

$$r=(x^2+z^2)^{1/2},$$

$$K_z^P=\frac{m_p m_o}{R}, \quad K_x^P=-\frac{iqm_o}{R},$$

$$K_z^S=-\frac{2m_p q^2}{R}, \quad K_x^S=\frac{2iqm_p m_s}{R},$$

$$R=m_0^2-4q^2m_p m_s \quad \text{(Rayleigh function)},$$

$$m_p=(q^2+1)^{1/2}, \quad m_s=(q^2+\gamma^2)^{1/2},$$

$$m_o=2q^2+\gamma^2, \quad \gamma=c_p/c_s,$$

$$t_{hw}=\frac{r}{c_p}[(\gamma^2-1)^{1/2}\cos(\theta)+\sin(\theta)],$$

$$t_s=r/c_s, \quad t_p=r/c_p$$

In these expressions the q value should be substituted by the one that corresponds in each case

$$q_p=\frac{c_p}{r^2}[itx+z(t^2-t_p^2)^{1/2}], \tag{A5}$$

$$q_s=\frac{c_p}{r^2}[itx+z(t^2-t_s^2)^{1/2}], \tag{A6}$$

$$q_{hw}=\frac{ic_p}{r^2}[tx-z(t_s^2-t^2)^{1/2}]. \tag{A7}$$

The directivity patterns for the longitudinal and lateral components correspond to the evaluation of Eqs. (A2) and (A3) as $t\to t_p^+$ and $t\to t_s^+$, respectively. The results are as follows:

$$G_z^P(t\to t_p^+)\cong\frac{1}{\pi\mu}\,\frac{N_{PZ}(\theta)}{D_P(\theta)}\,T_p^{-1}, \tag{A8}$$

$$G_x^P(t\to t_p^+)\cong\frac{1}{\pi\mu}\,\frac{N_{PX}(\theta)}{D_P(\theta)}\,T_p^{-1}, \tag{A9}$$

$$T_p=(t^2-t_p^2)^{1/2},$$

$$N_{PZ}(\theta)=\cos^2(\theta)[\gamma^2-2\sin^2(\theta)],$$

$$N_{PX}(\theta)=\sin(\theta)\cos(\theta)[\gamma^2-2\sin^2(\theta)],$$

$$D_P(\theta)=[\gamma^2-2\sin^2(\theta)]^2+4\sin^2(\theta)\times\cos(\theta)[\gamma^2-\sin^2(\theta)]^{1/2},$$

$$G_z^S(t\to t_s^+)\cong\frac{1}{\gamma\pi\mu}\,\frac{N_{SZ}(\theta)}{D_S(\theta)}\,T_s^{-1}, \tag{A10}$$

$$G_x^S(t\to t_s^+)\cong\frac{1}{\gamma\pi\mu}\,\frac{N_{SX}(\theta)}{D_S(\theta)}\,T_s^{-1}, \tag{A11}$$

$$T_s=(t^2-t_s^2)^{1/2},$$

$$N_{SZ}=\begin{cases}-8\sin^4(\theta)\cos^2(\theta)[\sin^2(\theta)-\gamma^{-2}] & \theta>\theta_c\\ -2\sin^2(\theta)\cos(\theta)[\gamma^{-2}-\sin^2(\theta)]^{1/2} & \theta<\theta_c,\end{cases}$$

$$N_{SX}=\begin{cases}-8\sin^3(\theta)\cos^3(\theta)[\sin^2(\theta)-\gamma^{-2}] & \theta>\theta_c\\ -2\sin(\theta)\cos^2(\theta)[\gamma^{-2}-\sin^2(\theta)]^{1/2} & \theta<\theta_c,\end{cases}$$

$$D_S=\begin{cases}[1-2\sin^2(\theta)]^4+16\sin^4(\theta)\cos^2(\theta)\times[\sin^2(\theta)-\gamma^{-2}] & \theta>\theta_c,\\ [1-2\sin^2(\theta)]^2+4\sin^2(\theta)\cos(\theta)\times[\gamma^{-2}-\sin^2(\theta)]^{1/2} & \theta<\theta_c.\end{cases}$$

[1] Y. Yamakoshi, J. Sato, and T. Sato, "Ultrasonic imaging of internal vibration of soft tissue under forced vibration," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37(2)**, 45–53 (1990).

[2] K. J. Parker, S. R. Huang, R. A. Muslin, and R. M. Lerner, "Tissue response to mechanical vibrations for sonoelasticity imaging," Ultrasound Med. Biol. **16(3)**, 241–246 (1990).

[3] N. Masuda, T. Tsujita, T. Ebuchi, and Y. Yamakoshi, ''Low-frequency elastic wave imaging by adaptive combination of fundamental and tissue harmonic ultrasound waves,'' Jpn. J. Appl. Phys., Part 1 **42**, 3271–3275 (2003).
[4] M. Jeong and S. Kwon, ''Tissue stiffness imaging method using temporal variation of ultrasound speckle pattern,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50(4)**, 457–460 (2003).
[5] S. Catheline, F. Wu, and M. Fink, ''A solution to diffraction biases in sonoelasticity: The acoustic impulse technique,'' J. Acoust. Soc. Am. **105(5)**, 2941–2950 (1999).
[6] L. Gao, K. J. Parker, and S. K. Alam, ''Sonoelasticity imaging: Theory and experimental verification,'' J. Acoust. Soc. Am. **97(6)**, 3875–3886 (1995).
[7] J. Ophir, S. K. Alam, B. Garra, F. Kallel, E. Konofagou, T. Krouskop, and T. Varghese, ''Elastography: Ultrasonic estimation and imaging of the elastic properties of tissues,'' Proc. Inst. Mech. Eng., Part H: J. Eng. Med. **213**, 203–233 (1999).
[8] E. J. Chen, J. Navakofski, W. K. Jenkins, and W. D. O'Brien, ''Young's modulus measurements of soft tissues with application to elasticity imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43(1)**, 191–194 (1996).
[9] L. Sandrin, M. Tanter, S. Catheline, and M. Fink, ''Shear modulus imaging with 2D transient elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49(4)**, 426–435 (2002).
[10] L. Sandrin, D. Casserau, and M. Fink, ''The role of the coupling term in transient elastography,'' J. Acoust. Soc. Am. **115(1)**, 73–83 (2004).
[11] A. Lhémery, ''A model for the transient ultrasonic field radiated by an arbitrary loading in a solid,'' J. Acoust. Soc. Am. **96(6)**, 3776–3786 (1994).
[12] D. Gridin, ''High frequency asymptotic description of head waves and boundary layers surrounding the critical rays in an elastic half-space,'' J. Acoust. Soc. Am. **104(3)**, 1188–1197 (1998).
[13] K. Aki and P. Richards, *Quantitative Seismology, Theory and Methods*, W. H. Freeman, San Francisco (1980).
[14] H. Djelouah and J. C. Baboux, ''Transient ultrasonic field radiated by a circular transducer in a solid medium,'' J. Acoust. Soc. Am. **92(5)**, 2932–2941 (1992).
[15] J. Bai, C. Ding, J. Luo, and P. He, ''Estimation and reduction of decorrelation effect due to tissue lateral displacement in elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49(5)**, 541–549 (2002).
[16] M. Bilgen and M. Insana, ''Deformation models and correlation analysis in elastography,'' J. Acoust. Soc. Am. **99(5)**, 3212–3224 (1996).
[17] D. C. Gakenheimer and J. Miklowitz, ''Transient excitation of an elastic half space by a point load traveling on the surface,'' J. Appl. Mech. **91**, 505–515 (1969).
[18] Z. Wu, L. S. Taylor, D. J. Rubens, and K. J. Parker, ''Shear wave focusing for three-dimensional sonoelastography,'' J. Acoust. Soc. Am. **111(1)**, 439–445 (2002).
[19] D. Evans and W. McDicken, *Doppler Ultrasound. Physics, Instrumentation and Signal Processing*, 2nd ed. (Wiley, New York, 2000).
[20] S. Catheline, ''Interferometrie-speckle ultrasonore: Application à la mesure d'elasticité,'' Thèse de doctoract de l'université Paris VII (Ultrasonic speckle interferometry: Application to elasticity measurement, Doctorate thesis of Paris VII University) (1998).
[21] W. F. Walker and G. E. Trahey, ''A fundamental limit on the accuracy of speckle signal alignment,'' IEEE 1994 Ultrasonics Symposium, 1787–1791.
[22] S. Catheline, J. L. Thomas, F. Wu, and M. Fink, ''Diffraction field of a low frequency vibrator in soft tissues using transient elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46(4)**, 1013–1019 (1999).
[23] E. Dieulesaint and D. Royer, *Elastic Waves in Solids* (Wiley, New York, 1981).
[24] J. Bercoff, S. Chaffai, M. Tanter, L. Sandrin, S. Catheline, M. Fink, J. L. Gennisson, and M. Meunier, ''*In vivo* breast tumor detection using transient elastography,'' Ultrasound Med. Biol. **29(10)**, 1387–1396 (2003).

Fluctuations in diffuse field–field correlations and the emergence of the Green's function in open systems

Richard L. Weaver[a] and Oleg I. Lobkis
Department of Theoretical and Applied Mechanics, University of Illinois at Urbana—Champaign, 216 Talbot Lab, 104 S. Wright Street, Urbana, Illinois 61801

(Received 1 December 2004; revised 14 February 2005; accepted 10 March 2005)

Recent intense interest in diffuse field correlation functions, with applications to passive imaging in underwater acoustics and seismology, has raised questions about the degree with which a retrieved waveform can be expected to conform to the Green's function, and in particular the degree with which a ray arrival may be discerned. On considering a simple scalar wave model consisting of fields with distributed random sources, the difffuse field-field correlation function R is defined as a sum of correlation integrals, one for each of the many distinct distributed sources. It is then shown that this ensemble of fields has a correlation function with expectation $\langle R \rangle$ equal to the Green's function. This model also lends itself to calculations of the variance of R, and thus to estimates of the degree to which an R calculated using finite amounts of data will conform to the Green's function. The model predicts that such conformation is strongest at low frequencies. Ray arrivals are detectable if sufficient data have been collected, but the amount of data needed scales in three dimensions with the square of the source–receiver separation, and the square of the frequency. Applications to seismology are discussed.
[DOI: 10.1121/1.1898683]

PACS numbers: 43.20.Gp, 43.40.Qi, 43.40.Ph [JJM] Pages: 3432–3439

I. INTRODUCTION

The recent proofs and demonstrations that diffuse fields have field–field correlation functions essentially equal to the Green's function[1–5] have been widely noted.[6–14] The potential for passive imaging is intriguing, especially in seismology where well-controlled sources are awkward, and in (notoriously noisy) ocean acoustics. The idea is not entirely new. Helioseismology[15] correlates acoustic fluctuations in the sun to determine the speeds of such waves. Claerbout and Cole, and Claerbout and Rickett[15] have speculated that correlations of ambient seismic waves would reveal geologic structures. Cole's attempts to verify and apply this were unsuccessful. Fehler *et al.*,[13] Wapenaar,[16] and Campillo and co-workers[6,7] have explored applications in seismology more recently, with promising results. Figure 1 illustrates the technique. A diffuse field is detected by two receivers at points a and b. An occasional ray passes through both detectors, thus generating weak correlations between their signals. In particular, we expect the correlation function to show a ray ''arrival'' at times $\pm|a-b|/c$ corresponding to propagation time over the distance $|a-b|$ at speed c. A related technique called ''microtremor survey'' is used to measure soil properties.[17]

Norton *et al.*[18] have speculated in similar ways; they describe a method for analysis of ambient seismic noise to resolve deep structures, a method that they term ''time exposure acoustics.'' Close examination, however, shows that their technique is not equivalent to that of Refs. 1–17. Similarly, the technique termed ''acoustic daylight''[19] bears some similarities, but is not equivalent. Nor are the field–field correlation functions of Refs. 1–17 equivalent to the better-known process of correlating a noisy source function with a noisy response (e.g., Ref. 20) to construct impulse responses.

The proof by Weaver and Lobkis[2] was based on a modal expansion for the diffuse field. They showed that convergence is faster if time integrations can be carried out over ranges comparable to the Heisenberg time t_H at which modes can be resolved and/or if sources are distributed over all space. This was perhaps of small concern in the corresponding laboratory demonstrations conducted in finite bodies.[2] But, in open systems where t_H is infinite the potential restrictions are problematic. Recently, the proof was extended to open homogeneous[11,12] and open heterogeneous systems.[21]

Both in laboratory demonstrations,[2,9,10] and in attempts to apply the method to seismic waves, correlation functions have in practice imperfectly matched the Green function. Lobkis and Weaver,[2] using diffuse ultrasonic waves in a finite reverberant system, found correlation functions which showed one or two ray arrivals with distorted waveforms, on a background of random fluctuations. They observed surface waves and reflections from nearby interfaces in finite bodies, but only after extensive averaging, in time and over a small number of distinct sources. Even then, they observed convergence to be poor at higher frequencies. Using thermal phonons,[1,3,4] and analyzing much more data than in Ref. 2, they resolved more arrivals. Roux *et al.*[14] showed direct arrivals and some bottom and surface reflections in the correlations of ocean shipping noise. Derode *et al.*[9] studied ultrasonic correlations in a multiply scattering water tank. They employed a carefully chosen set of diffuse field sources and obtained an excellent agreement between directly obtained waveforms and correlations, with a quality comparable to that seen using thermal fluctuations.[1,3,4] In seismology,

[a]Electronic mail: r-weaver@uiuc.edu

 0001-4966/2005/117(6)/3432/8/$22.50

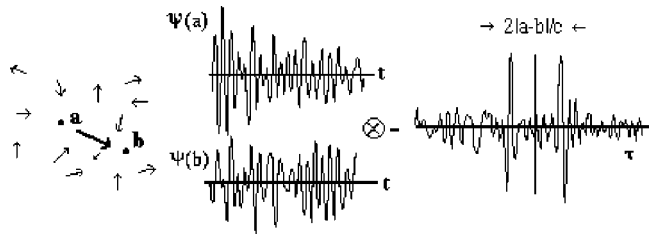


FIG. 1. Detectors at positions a and b are immersed in a diffuse wave field. An occasional ray, for example the one in bold, passes through both detectors, thus generating weak correlations between the noisy signals $\psi(a,t)$ and $\psi(b,t)$, reported by the receivers. The correlation function $C_{ab}(\tau)$ shows peaks at the ray arrival times $\tau=\pm|a-b|/c$. If background fluctuations are too strong, as they will be if averaging has been insufficient, they will obscure the ray arrivals.

Campillo and Paul[6] observed low-frequency Rayleigh and Love waves in the correlations of seismic codas on a regional scale, and Shapiro and Campillo observed[7] very low-frequency surface waves in correlations of diffuse seismic noise on a continental scale. They did not observe other wave types, or higher frequencies. Shapiro *et al.*[22] and Sabra *et al.*[23] applied these ideas to correlations of low-frequency seismic noise and generated unprecedentedly high-resolution seismic wave velocity maps. They too observed mostly surface waves. These papers have thus implicitly shown our need for theoretical understanding of the quality and *rate* of convergence. How much averaging is necessary? How large will the residual fluctuations be? What parts of the Green function may we expect to retrieve without ambiguity? It is towards these questions that this work is addressed.

Here, we confine attention to open systems, where most applications appear to reside. In such systems arguments that depend on time integrations over the scale of the Heisenberg time cannot be admitted. It remains acceptable to use discrete modal expansions though, as sums remain discrete as long as the system is finite, although it can be as large as desired.

In finite systems, ergodically filled by diffuse waves, it is convenient to take a diffuse field in the form of a modal expansion with uncorrelated modal amplitudes with equal mean squares. Thus, the diffuse field fills the entire volume uniformly. In open systems this definition is not satisfactory; it is clear for example that neither the ocean or the earth is fully diffuse in that sense. Instead, we imagine our region of interest to have a diffuse field that is local. As in Ref. 21, this is done by taking the region to be insonified by an uncorrelated superposition of incident waves (i.e., an incident field that conforms to the room acoustics notion of a diffuse field). It is recognized, as in Ref. 21, that such a field is conveniently represented as generated by a process spread over all space with delta-correlated spatial statistics and nonstochastic spectrum.

Such an analytical model is introduced in Sec. II, after which the field–field correlation functions are shown, in Sec. III, to have their familiar mean expectation, equal to the Green's function of the medium. The mean-square correlation function is derived in the following section, and the corresponding variance is compared with the square of the mean. It is found that the variance is small compared with the square of the mean if averaging is sufficient. The precise meaning of "sufficient" is described mathematically. Implications for the seismic case are discussed.

II. MODEL

Consider the modal sum quantity

$$C_{ab}(t)\equiv\sum_n u^n(a)u^n(b)\cos(\omega_n t), \tag{1}$$

defined in the sense of distributions for all t. Here, $u^n(a)=u_a^n$ is the nth mode, evaluated at position a. The natural frequency of the mode is ω_n. C is a time-symmetrized version of the velocity Green's function[24]

$$\dot{G}_{ab}(t)\equiv\Theta(t)\sum_n u^n(a)u^n(b)\cos(\omega_n t), \tag{2}$$

as follows from the identity

$$C_{ab}(t)\equiv\dot{G}_{ab}(t)+\dot{G}_{ab}(-t). \tag{3}$$

In the special case of an unbounded scalar homogeneous medium, G is the solution to

$$[-c^2\nabla^2+\partial^2/\partial t^2]G_{xy}(t)=\delta(t)\,\delta(x-y), \tag{4}$$

which in three dimensions has the solution

$$G_{xy}(t)=\frac{\delta(t-|x-y|/c)}{4\pi|x-y|c^2}. \tag{5}$$

We further define a time-varying signal $\Psi_a^\sigma(t)$ as the response at position a to a spatially distributed source $\dot{s}_x^\sigma(t)$, after filtering by a receiver function $r(t)$

$$\Psi_a^\sigma(t)=\int dx\int d\phi\,d\alpha\dot{s}_x^\sigma(\phi)r(\alpha)G_{ax}(t-\phi-\alpha)$$
$$=\int dx\int d\phi\,d\alpha s_x^\sigma(\phi)r(\alpha)\dot{G}_{ax}(t-\phi-\alpha). \tag{6}$$

A finite set of distinct random sources $s_x^\sigma(t)$ $(\sigma=1,2,...,N)$ with Gaussian statistics

$$\langle s_x^\sigma\rangle=0;\quad \langle s_x^\sigma(t)s_{x'}^{\sigma'}(t')\rangle=\delta(x-x')s(t)s(t')\,\delta_{\sigma\sigma'} \tag{7}$$

$$\langle s_x^\sigma(t)s_{x'}^{\sigma'}(t')s_y^{\sigma''}(t'')s_{y'}^{\sigma'''}(t''')\rangle$$
$$=\langle s_x^\sigma(t)s_{x'}^{\sigma'}(t')\rangle\langle s_y^{\sigma''}(t'')s_{y'}^{\sigma'''}(t''')\rangle$$
$$+\langle s_x^\sigma(t)s_y^{\sigma''}(t'')\rangle\langle s_{x'}^{\sigma'}(t')s_{y'}^{\sigma'''}(t''')\rangle$$
$$+\langle s_x^\sigma(t)s_{y'}^{\sigma'''}(t''')\rangle\langle s_{x'}^{\sigma'}(t')s_y^{\sigma''}(t'')\rangle, \tag{8}$$

will be applied to the system. A typical sample $s_x(t)$ from the ensemble may be pictured as a white-noise process in space, with nonstochastic time dependence $s(t)$. Such a source is unrealistic in practice. It is used here because it is a convenient means by which the region around points a and b may be insonified by a classically Gaussian diffuse field.[21]

The correlation R is defined as an average over N samples from the ensemble, of the field–field correlation function as evaluated by integrating over a (long) window $W(t)$

$$R_{ab}(\tau)=\frac{1}{N}\sum_{\sigma}\int dt\, W(t)\Psi_a^{\sigma}(t)\Psi_b^{\sigma}(t+\tau). \tag{9}$$

It is this quantity that is expected to converge to the Green's function. In practice the fields Ψ suffer from absorption, and/or are weak when sources are distant. A naive use of raw signals Ψ in the above therefore risks a nonuniform weighting of the available useful information. Lobkis and Weaver[2] compensated for absorption by augmenting late time signals with a frequency-dependent gain. Larose *et al.*[9] employed the much simpler process of replacing Ψ with its one-bit version, thereby balancing the contributions from different amplitudes. Campillo and co-workers[6,7] carried out similar processing on their seismic signals to maximize useful information. The present analysis considers no such nonlinear processing, the assumed source statistics rendering it unnecessary. It is presumed that, for comparisons with practice, the effective duration of the experimentalist's window W may be estimated.

III. THE MEAN CORRELATION FUNCTION $\langle R\rangle$

It is convenient to take the window W to be rectangular

$$W=1 \quad \text{if } t_o-T/2<t<t_o+T/2; \quad W=0 \quad \text{otherwise}, \tag{10}$$

where $T=\int W(t)dt$ is the amount of time over which signal Ψ is collected. The center time t_o is understood to be large enough that $W(t=0)$ is zero; $t_o>T/2$. T and t_o are also taken to be much greater than the correlation times τ of interest. Under these conditions, we may replace $\dot{G}(t)$ with $C(t)$ in (6). Then

$$\begin{aligned}R_{ab}(\tau)&=\frac{1}{N}\sum_{\sigma}\int dt\, W(t)\Psi_a^{\sigma}(t)\Psi_b^{\sigma}(t+\tau)\\&=\frac{1}{N}\sum_{\sigma}\int W(t)dt\,dx\,dy\,d\phi\,d\psi\,d\alpha\,d\beta\\&\quad\times s_x^{\sigma}(\phi)s_y^{\sigma}(\psi)r(\alpha)r(\beta)C_{ax}(t-\phi-\alpha)\\&\quad\times C_{by}(t+\tau-\psi-\beta).\end{aligned} \tag{11}$$

For simplicity, it has been assumed that the receivers at a and b have identical transfer functions $r(t)$.

The expectation of R is obtained by substituting for $\langle ss\rangle$ from (7)

$$\begin{aligned}\langle R_{ab}(\tau)\rangle=\int dtW(t)dx\,d\phi\,d\psi\,d\alpha\,d\beta\,s(\phi)s(\psi)r(\alpha)\\\times r(\beta)C_{ax}(t-\phi-\alpha)C_{bx}(t+\tau-\psi-\beta).\end{aligned} \tag{12}$$

The integral over x may be evaluated by use of a Ward identity[25]

$$\int dx\, C_{ax}(t)C_{bx}(t')=\frac{1}{2}[C_{ab}(t+t')+C_{ab}(t-t')]. \tag{13}$$

The identity is proved by substituting for C in terms of the modes, and invoking their orthonormality. Alternatively, it may be proved by reciprocity arguments as in Refs. 16, 21. Thus

$$\begin{aligned}\langle R_{ab}(\tau)\rangle&=\int dt\, W(t)d\phi\,d\psi\,d\alpha\,d\beta\; r(\alpha)r(\beta)s(\phi)s(\psi)\\&\quad\times\frac{1}{2}[C_{ab}(2t+\tau-\phi-\psi-\alpha-\beta)\\&\quad+C_{ab}(\tau+\phi-\psi+\alpha-\beta)].\end{aligned} \tag{14}$$

The α, β, ϕ, and ψ integrations filter these Cs. On denoting such quadruply filtered Cs as $C^{=}$ and $C^{\pm}$, the above becomes

$$\langle R_{ab}(\tau)\rangle=\frac{1}{2}\int dt\, W(t)[C_{ab}^{=}(2t+\tau)+C_{ab}^{\pm}(\tau)]. \tag{15}$$

The superscript $=$ denotes two filterings by s, and two by r. The superscript $\pm$ denotes filtering by s and by r, and an additional filtering by the time-reversed s and the time-reversed r.

On integration with respect to t, the first term is essentially zero. (We assume the sources and receivers have no amplitude at zero frequency.) The second term then gives

$$\langle R_{ab}(\tau)\rangle=\frac{T}{2}C_{ab}^{\pm}(\tau), \tag{16}$$

thus once again establishing the now-familiar identification of the expectation of field–field correlations R_{ab} with the (time derivative of the time-symmetrized) Green's function C_{ab}.

In order to be able to compare $\langle R\rangle$ with the expected level of R's fluctuations, it is necessary to estimate its amplitude. We note that

$$\langle R\rangle=TC_{ab}^{\pm}/2=(T/2)C_{ab}\otimes f, \tag{17}$$

where f is the product of the filters and $\otimes$ represents a convolution.

$$\tilde{f}(\omega)=|\tilde{r}(\omega)|^2|\tilde{s}(\omega)|^2. \tag{18}$$

For the purposes of illustration $f(t)$ is taken to be a Gaussian tone burst with central frequency $\overline{\omega}$ and duration $\Delta\gg 1/\overline{\omega}$

$$f(t)=\exp(-t^2/2\Delta^2)\cos(\overline{\omega}t). \tag{19}$$

Thus

$$\begin{aligned}\tilde{f}(\omega)=\frac{\sqrt{2\pi}\Delta}{2}[\exp\{-(\omega-\overline{\omega})^2\Delta^2/2\}\\+\exp\{-(\omega+\overline{\omega})^2\Delta^2/2\}].\end{aligned} \tag{20}$$

For the special case of the first ray arrival between a and b in a locally homogeneous medium, G is given by Eq. (5). Thus, $\langle R\rangle$ is

$$\langle R\rangle = TC_{ab}^{\pm}/2$$
$$= (T/2)C_{ab}\otimes f \approx (T/8\pi rc^2)$$
$$\times \exp(-(t-r/c)^2/2\Delta^2)\overline{\omega}\sin\{\overline{\omega}(t-r/c)\} + (t\leftrightarrow -t), \tag{21}$$

where $r=|a-b|$. This is a Gaussian tone burst with peak amplitude $T\overline{\omega}/8\pi rc^2$.

IV. MEAN-SQUARE ESTIMATE $\langle R^2\rangle$

We now investigate the fluctuations of R. The square of the correlation estimate [Eq. (9)] is

$$R_{ab}^2(\tau) = N^{-2}\sum_{\sigma}\sum_{\sigma'}\int W(t)W(t')dt\,dt'\Psi_a^{\sigma}(t)\Psi_b^{\sigma}(t+\tau)\Psi_a^{\sigma'}(t')\Psi_b^{\sigma'}(t'+\tau)$$
$$= N^{-2}\sum_{\sigma}\sum_{\sigma'}\int W(t)W(t')dt\,dt'd\phi\,d\phi'd\psi\,d\psi'd\alpha\,d\alpha'd\beta\,d\beta'dx\,dx'dy\,dy'r(\alpha)r(\beta)r(\alpha')r(\beta')$$
$$\times s_x^{\sigma}(\phi)s_y^{\sigma}(\psi)s_{x'}^{\sigma'}(\phi')s_{y'}^{\sigma'}(\psi')C_{ax}(t-\phi-\alpha)C_{by}(t+\tau-\psi-\beta)C_{ax'}(t'-\phi'-\alpha')C_{by'}(t'+\tau-\psi'-\beta'). \tag{22}$$

The expectation of this is obtained by substituting for $\langle ssss\rangle$ from Eq. (8)

$$\langle R_{ab}^2(\tau)\rangle = \langle R_{ab}(\tau)\rangle^2 + \mathrm{var}_{ab}(\tau), \tag{23}$$

where var is the variance and is given by, after further invocations of the identity (13)

$$\mathrm{var}_{ab}(\tau) = \frac{1}{N}\int WWdt\,dt'd\phi\,d\psi\,d\phi'd\psi'd\alpha\,d\alpha'd\beta\,d\beta'r(\alpha)r(\beta)r(\alpha')r(\beta')s(\phi)s(\psi)s(\phi')s(\psi')$$
$$\times\left\{\begin{array}{l}[C_{aa}(t+t'-\phi-\phi'-\alpha-\alpha')+C_{aa}(t-t'-\phi+\phi'-\alpha+\alpha')]\\ \quad\times[C_{bb}(t+t'+2\tau-\psi-\psi'-\beta-\beta')+C_{bb}(t-t'-\psi+\psi'-\beta+\beta')]\\ +[C_{ab}(t+t'+\tau-\phi-\psi'-\alpha-\beta')+C_{ab}(t-t'-\phi+\psi'-\tau-\alpha+\beta')]\\ \quad\times[C_{ab}(t+t'+\tau-\psi-\phi'-\alpha'-\beta)+C_{ab}(t-t'+\tau-\psi+\phi'-\beta+\alpha')]\end{array}\right\}, \tag{24}$$

or

$$\mathrm{var}_{ab}(\tau) = \frac{1}{4N}\int W(t)W(t')dt\,dt'\left\{\begin{array}{l}[C_{aa}^{=}(t+t')+C_{aa}^{\pm}(t-t')][C_{bb}^{=}(t+t'+2\tau)+C_{bb}^{\pm}(t-t')]\\ \quad+[C_{ab}^{=}(t+t'+\tau)+C_{ab}^{\pm}(t-t'-\tau)][C_{ab}^{=}(t+t'+\tau)+C_{ab}^{\pm}(t-t'+\tau)]\end{array}\right\}. \tag{25}$$

The integrations over t and t' are best performed by changing variables

$$\xi = t+t'-2t_o; \quad \eta = t-t', \tag{26}$$

such that

$$\int \ldots W(t)W(t')dt\,dt' = \frac{1}{2}\int_{\xi=-T}^{T}\int_{\eta=|\xi|-T}^{T-|\xi|}\ldots d\eta\,d\xi = \frac{1}{2}\int_{\eta=-T}^{T}\int_{\xi=|\eta|-T}^{T-|\eta|}\ldots d\xi\,d\eta \tag{27}$$

whereupon

$$\mathrm{var}_{ab}(\tau) = \frac{1}{8N}\int_{\xi=-T}^{T}\int_{\eta=|\xi|-T}^{T-|\xi|}d\eta\,d\xi\left\{\begin{array}{l}[C_{aa}^{=}(\xi+2t_o)+C_{aa}^{\pm}(\eta)][C_{bb}^{=}(\xi+2t_o+2\tau)+C_{bb}^{\pm}(\eta)]\\ \quad+[C_{ab}^{=}(\xi+2t_o+\tau)+C_{ab}^{\pm}(\eta-\tau)][C_{ab}^{=}(\xi+2t_o+\tau)+C_{ab}^{\pm}(\eta+\tau)]\end{array}\right\}. \tag{28}$$

The cross terms integrate to zero. Typical among these terms is

$$\int d\xi\,d\eta C_{aa}^{=}(\xi+2t_o)C_{bb}^{\pm}(\eta).$$

The ξ integration vanishes, by virtue of the lack of low-frequency components in r and s. The η integration may not vanish (see below) but is in any case small, and multiplied by the vanishing ξ integration. We therefore neglect the cross terms and summarize the others

$$\mathrm{var}_{ab}(\tau)=\frac{1}{8N}\int\int d\xi\, d\eta\left\{\begin{array}{l} C^{=}_{aa}(\xi+2t_o)C^{=}_{bb}(\xi+2t_o+2\tau)+C^{\pm}_{aa}(\eta)C^{\pm}_{bb}(\eta) \\ \quad +C^{=}_{ab}(\xi+2t_o+\tau)C^{=}_{ab}(\xi+2t_o+\tau)+C^{\pm}_{ab}(\eta-\tau)C^{\pm}_{ab}(\eta+\tau)\end{array}\right\}$$

$$=\int_{-T}^{T}(T-|\xi|)d\xi C^{=}_{aa}(\xi+2t_o)C^{=}_{bb}(\xi+2t_o+2\tau)/4N+\int_{-T}^{T}(T-|\xi|)d\xi C^{=}_{ab}(\xi+2t_o+\tau)C^{=}_{ab}(\xi+2t_o+\tau)/4N$$

$$+\int_{-T}^{T}(T-|\eta|)d\eta\; C^{\pm}_{aa}(\eta)C^{\pm}_{bb}(\eta)/4N+\int_{-T}^{T}(T-|\eta|)d\eta\; C^{\pm}_{ab}(\eta-\tau)C^{\pm}_{ab}(\eta+\tau)/4N$$

$$=I+II+III+IV. \tag{29}$$

The first term I integrates to approximately zero, by virtue of the lack of correlations between C_{aa} and C_{bb}

$I=0$.

The second term is $T^2/4N$ times an average of the square of C_{ab}. This average is performed over the time interval from $2t_o-T$ to $2t_o+T$, with a triangular weighting that emphasizes the times near $2t_o$.

$$II=[T^2/4N]\overline{C^{=2}_{ab}}^{\text{times centered near } 2t_o}. \tag{30}$$

The third term III might appear to vanish, by means of the argument that was applied to the first term I. But, there is a coherent contribution in the vicinity of $\eta=0$. Indeed, $C_{pp}(\eta)$ (before filtering) has a delta-function-like behavior $\delta(\eta)$ near $\eta=0$. We note

$$C^{\pm}_{pp}(\eta)=\sum_n\;(u^n_p)^2|\tilde{s}(\omega_n)|^2|\tilde{r}(\omega_n)|^2\cos(\omega_n\eta). \tag{31}$$

So

$$\int_{-T}^{T}d\eta\;(T-|\eta|)C^{\pm}_{aa}(\eta)C^{\pm}_{bb}(\eta)/4N=\frac{1}{4N}\sum_{n,m}\;(u^n_a)^2(u^m_b)^2|\tilde{s}(\omega_n)|^2|\tilde{s}(\omega_m)|^2|\tilde{r}(\omega_n)|^2|\tilde{r}(\omega_m)|^2$$

$$\times\int_{-T}^{T}d\eta(T-|\eta|)\cos(\omega_m\eta)\cos(\omega_n\eta) \tag{32}$$

$$=\frac{1}{4N}\sum_{n,m}\;(u^n_a)^2(u^m_b)^2|\tilde{s}(\omega_n)|^2|\tilde{s}(\omega_m)|^2|\tilde{r}(\omega_n)|^2|\tilde{r}(\omega_m)|^2$$

$$\times\int_{-T}^{T}d\eta(T-|\eta|)\,\frac{1}{2}\{\cos((\omega_m+\omega_n)\eta)+\cos[(\omega_m-\omega_n)\eta]\}. \tag{33}$$

The first term vanishes on performing the integration $d\eta$. The second gives

$$=\frac{1}{4N}\sum_{n,m}\;(u^n_p)^2(u^m_p)^2|\tilde{s}(\omega_n)|^2|\tilde{s}(\omega_m)|^2|\tilde{r}(\omega_n)|^2|$$

$$\times\tilde{r}(\omega_m)|^2\frac{1-\cos\{T(\omega_n-\omega_m)\}}{(\omega_n-\omega_m)^2}. \tag{34}$$

The modal sums are replaced with integrals

$$=\sum_{n,m}\cdots(u^n_p)^2(u^m_p)^2\Rightarrow\frac{1}{V^2}\int\int\cdots D(\omega)D(\omega')d\omega\, d\omega', \tag{35}$$

where V is system volume and D is modal density. The integration is dominated by points ω, ω' that are close, within $1/T$, and so the smoothly varying filters r and s and the modal density D may be replaced with their values at the mean frequency $\omega_c=(\omega+\omega')/2$. Thus

$$III=\frac{1}{4N}\int_{-T}^{T}d\eta(T-|\eta|)C^{\pm}_{aa}(\eta)C^{\pm}_{bb}(\eta)$$

$$=\frac{1}{4NV^2}\int D^2(\omega_c)|\tilde{s}(\omega_c)|^4|\tilde{r}(\omega_c)|^4d\omega_c$$

$$\times\int d\omega_d\frac{1-\cos\{T\omega_d\}}{\omega_d^2}$$

$$=\frac{\pi T}{4NV^2}\int D^2(\omega_c)|\tilde{s}(\omega_c)|^4|\tilde{r}(\omega_c)|^4d\omega_c. \tag{36}$$

The fourth term of (29) is

$$IV=\int_{-T}^{T}(T-|\eta|)d\eta\,C^{\pm}_{ab}(\eta-\tau)C^{\pm}_{ab}(\eta+\tau)/4N. \tag{37}$$

At $\tau=0$, this is $T^2/4N$ times a time-averaged C^2_{ab}, averaged over a time period $(-T,+T)$ centered on zero with a triangular weighting. At $\tau\neq0$, it is the same, multiplied by the

factor $d(\tau)=As\otimes s^{TR}\otimes r\otimes r^{TR}$, with A such that $d(0)=1$ and 'TR' indicating time-reversal.

In sum, var is

$$\mathrm{var}_{ab}(\tau)=\frac{T^2}{4N}[\overline{C_{ab}^{=2}}\text{ times centered at }2t_o + d(\tau)\overline{C_{ab}^{\pm 2}}\text{ times centered at }0] + \frac{\pi T}{4NV^2}\int_0^\infty D^2(\omega_c)|\tilde{s}(\omega_c)|^4|\tilde{r}(\omega_c)|^4 d\omega_c . \tag{38}$$

The first term may be estimated by integrating it over all a, and dividing by a characteristic volume, the volume V_{2to}, which the field fills by time $2t_o$

$$\overline{C_{ab}^{=}(t)^2}\equiv\frac{1}{V_{2t_o}}\int da\sum_n\sum_m u^n(a)u^n(b)u^m \times(a)u^m(b)\,\overline{\cos(\omega_n t)\cos(\omega_m t)} = \frac{1}{V_{2t_o}}\sum_m \overline{\cos^2(\omega_n t)}u^n(b)^2 = \frac{1}{2V_{2t_o}}\int d\omega\,\rho(\omega)|\tilde{s}(\omega)|^4|\tilde{r}(\omega)|^4,$$

where $\rho=D/V$, the modal density per volume. If a and b are close, and if the field diffuses in that region with diffusivity $\mathcal{D}$, this volume is of order $(2\mathcal{D}t_o)^{3/2}$.

Similarly, the second term may be estimated by

$$\overline{C_{ab}^{\pm 2}}\text{ times centered at }0=\frac{1}{2V_o}\int d\omega\,\rho(\omega)|\tilde{s}(\omega)|^4|\tilde{r}(\omega)|^4, \tag{39}$$

where V_o is the volume filled by C_{ab} at times of the order of $T/2$.

$$\mathrm{var}\sim\frac{T^2}{8N}\left\{\int d\omega\,\rho(\omega)|\tilde{s}(\omega)|^4|\tilde{r}(\omega)|^4\left(\frac{1}{V_{2t_o}}+\frac{d(\tau)}{V_o}\right) + \int\rho(\omega_c)^2|\tilde{s}(\omega_c)|^4|\tilde{r}(\omega_c)|^4 d\omega_c\left(\frac{2\pi}{T}\right)\right\}. \tag{40}$$

Recalling that $\langle R\rangle$ itself scales with T, we see that increased T decreases the contribution of the last term relative to $\langle R\rangle^2$, and does so explicitly. The contributions of the other terms to the relative variance fall with T also, because the effective volumes increase with T and t_o. That the variance drops like $1/T$ was predicted by Sneider.[11]

The third term dominates. The ratio of the third term to the others is of the order of the ratio t_H/T of the Heisenberg time $t_H=2\pi V_{2to}\rho$ of the volume V_{2to} filled by the field, to the observation time T. This ratio is large (otherwise there is Anderson localization). Inasmuch as the third term with $\rho\sim\omega^2$ is strongly dependent on the frequency passed by the filters s and r, it appears that high frequencies are the most problematic; their variances are large, so convergence is slower at high frequencies. This is consistent with observations.[2,7]

V. RESOLUTION OF A RAY ARRIVAL

If a ray arrival is to be detected in a construction R like (11), it must have amplitude greater than typical fluctuations $\sqrt{\mathrm{var}}$. To make such comparisons $f(t)$ is taken, as before, to be a Gaussian tone burst (19). The dominant term of var is therefore

$$\mathrm{var}\sim\frac{T^2}{8N}\frac{2\pi}{T}\rho(\bar{\omega})^2\left\{\int_0^\infty|\tilde{s}(\omega)|^4|\tilde{r}(\omega)|^4 d\omega\right\},$$

where

$$\tilde{f}(\omega)=|\tilde{s}(\omega)|^2|\tilde{r}(\omega)|^2 = \frac{\sqrt{2\pi}\Delta}{2}[\exp\{-(\omega-\bar{\omega})^2\Delta^2/2\} + \exp\{-(\omega+\bar{\omega})^2\Delta^2/2\}]. \tag{41}$$

So

$$\mathrm{var}\sim\frac{\pi^2 T}{8N}\Delta^2\rho(\bar{\omega})^2\left\{\int\exp[-(\omega-\bar{\omega})^2\Delta^2]d\omega\right\} = \frac{\pi^{5/2}T\Delta}{8N}\rho(\bar{\omega})^2 = \frac{\pi^{5/2}T\Delta}{32N}\frac{\bar{\omega}^4}{c^6\pi^4}. \tag{42}$$

This may be compared to the square of the peak amplitude of the first ray arrival $(T\bar{\omega}/8\pi rc^2)^2$, as derived following Eq. (21). The ratio is

$$\frac{1}{2\sqrt{\pi}}\frac{NT}{\Delta}\frac{c^2}{r^2\bar{\omega}^2}. \tag{43}$$

This ratio must be greater than unity if the arrival is to be detected. It is seen, therefore, that a ray arrival may be (barely) detected if the length NT of the data record from which the correlation is constructed times the half-power bandwidth HPBW$=2.2/\Delta$ of the frequency band used is sufficiently greater than the square of the nondimensional source/receiver separation, $r/\lambda=|a-b|/\lambda$. It is apparent that high frequencies, and low-amplitude ray arrivals, are the more difficult to resolve.

This calculation has been illustrated for the case of scalar waves in unbounded three-dimensional media. Its predictions should be relevant to correlations in underwater ultrasonics[8,9] and in ocean acoustics.[12,14] With a view towards applications in seismology, it will need to be adapted to the case of elastic waves in a half-space, and in particular to the surface wave arrivals seen in seismic noise correlations.[6,7,22,23] Without rederiving (43) for this case, we may roughly infer the form it would take for such waves. The variance will preserve its $\rho(\bar{\omega})^2\sim\bar{\omega}^4$ dependence. But, Rayleigh waves spread in two dimensions, diminishing asymptotically in amplitude as $[c/\bar{\omega}r]^{1/2}$. Thus, we expect (43) to scale differently with frequency and distance, the ratio now being $B(NT/\Delta)(c/r\bar{\omega})$ with an as yet undetermined numerical factor B. This prediction can be applied to the seismic data of Campillo and co-workers.[6,7] Shapiro and

Campillo[7] cross correlated 5 weeks of ambient seismic noise at stations in New Mexico and Missouri (a distance of 1500 km). They found that the direct Rayleigh wave between these stations could be observed clearly at all frequencies up to about 0.08 Hz. As judged by eye from their figures, their signal-to-noise ratios scale as predicted: proportional to the inverse of frequency. If the entire 5-week record is identified with NT, we estimate $NT=5$ weeks$=3\times10^6$ s. We furthermore take $r=1500$ km; $c=3.3$ km/s; and examine the record for $[0.035<f<0.08\,\mathrm{Hz}]$, i.e., $\bar{\omega}=0.36$ rad/s, HPBW$=0.045$ Hz$=0.28$ rad/s, so $\Delta=7.8$ s. The predicted signal-to-noise ratio $(NT/\Delta)(c/r\bar{\omega})$ is then about 2300. This compares poorly with the observed ratio (~4) of the square of the Rayleigh wave's peak amplitude to the mean square of the background [Fig. 1(b) of Ref. 7]. The probable inference is that their 5-week record is dominated by local and uncorrelated noise sources (e.g., wind, machinery, local microseisms) at each station; the effective amount of data is much less than 5 weeks. This could perhaps be modeled by modifying the assumed homogeneous white-noise source statistics of Eqs. (7) and (8).

Campillo and Paul[6] correlated coda signals, from 101 local earthquakes, between nearby stations in Mexico. Each coda had duration of a few hundred seconds. NT is therefore estimated at 4×10^4 s. They presented correlations between seismic stations PLIG and YAIG (69-km spacing) and showed good resolution of $\sim$0.12-Hz Rayleigh and Love waves. Taking $\Delta=8$ s and $c\sim3.5$ km/s, a predicted signal-to-noise ratio of $B(NT/\Delta)(c/r\bar{\omega})=282\ B$ is derived. It is therefore not surprising that their ray arrivals have substantially more ($\sqrt{282B}=17\sqrt{B}$) amplitude than the background at the same frequency. That they did not observe any ray arrivals at higher frequency is perhaps due, as they speculate, to excess absorption in the coda at those frequencies; the effective value of NT is too small at those frequencies.

VI. SUMMARY

Diffuse field correlations are attracting attention in underwater acoustics and in seismology with a view towards applications in passive imaging. Critical to successful retrieval of meaningful waveforms is an understanding of the rate at which such correlations will converge; how much averaging is necessary? What features can one expect to resolve? It has been shown here, using a simple model for a diffuse field in an open medium, that the variance of a band-limited field–field correlation function is proportional to the square of the central frequency and the first power of the signal duration captured. On comparing with the amplitude of expected ray arrivals in the correlation function, it is shown that the amount of data needed to resolve a ray arrival out of the fluctuations scales with the square of the frequency and the square of the distance $|a-b|$. This is qualitatively consistent with observations in the field and in the laboratory to date, but more measurements will be necessary before the predictions can be confirmed in detail.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation INT-0230312 and CMS-0201346. We thank Michel Campillo for helpful discussions.

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

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

[3] R. L. Weaver and O. I. Lobkis, "On the emergence of the Greens function in the correlations of a diffuse field; pulse-echo using thermal phonons," Ultrasonics **40**, 435–439 (2002).

[4] R. L. Weaver and O. I. Lobkis, "Elastic wave thermal fluctuations, ultrasonic waveforms by correlation of thermal phonons," J. Acoust. Soc. Am. **113**, 2611–2621 (2003).

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

[6] M. Campillo and A. Paul, "Long range correlations in the diffuse seismic coda," Science **299**(5606), 547–549 (2003).

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

[8] P. Roux and M. Fink, "Green's function estimation using secondary sources in a shallow water environment," J. Acoust. Soc. Am. **113**, 1406–1416 (2003).

[9] A. Derode, E. Larose, M. Campillo, and M. Fink, "How to estimate the Green's function of a heterogeneous medium between two passive sensors? Application to acoustic waves," Appl. Phys. Lett. **83**, 3054–3056 (2003). E. Larose, A. Derode, M. Campillo, and M. Fink, "Imaging from one-bit correlations of wideband diffuse wave fields," J. Appl. Phys. **95**, 8393–8399 (2004); A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from field–field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973 (2003).

[10] A. E. Malcolm, J. A. Scales, and B. A. van Tiggelen, "Retrieving the Green function from diffuse, equipartitioned waves," Phys. Rev. E **70**, 015601(R) (2004).

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

[12] P. Roux, K. Sabra, W. A. Kupperman, and A. Roux, "Ambient noise cross-correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79 (2005).

[13] K. G. Sabra, P. Roux, and W. A. Kuppeman, "Arrival-time structure of the time-averaged ambient noise cross-section fucntion in an oceanic waveguide," J. Acous. Soc. Am. **117**, 164–174 (2005).

[14] P. Roux and W. A. Kuperman, "Coherent waveforms from ocean noise," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).

[15] T. L. Duvall, S. M. Jefferies, J. W. Harvey, and M. A. Pomerantz, "Time-distance Helioseismology," Nature (London) **362**, 430 (1993); J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge **18**, 957–960 (1999).

[16] K. Wapenaar, "Synthesis of an inhomogeneous medium from its acoustic transmission response," Geophysics **68**, 1756–1759 (2003); K. Wapenaar, J. Thorbecke, and D. Draganov, "Relations between reflection and transmission responses of three-dimensional inhomogeneous media," Geophys. J. Int. **156**(2), 179–194 (2004).

[17] Hiroshi Okada "The microtremor survey method" Geophysical Monogr. Series no **12**, Soc Exploration Geophysicists, 2003; J. Loui, Bull. Seismol. Soc. Am. **91**, 347–364 (2001).

[18] S. J. Norton and I. J. Won, "Time exposure acoustics," IEEE Trans. Geosci. Remote Sens. **38**(3), 1337–1343 (2000); S. J. Norton and T. Vo-Dinh, "Diffraction tomographic imaging with photon density waves: An explicit solution," J. Opt. Soc. Am. A **15**(10), 2670–2677 (1998); W. E. Doll, S. J. Norton, L. J. Gray, and M. D. Morris, "Passive seismic imaging of scatterers," Oak Ridge National Lab, Oak Ridge, TN Tech. Rep. K/NSP-585, Oct. 1997.

[19] J. R. Potter, ''Acoustic imaging using ambient noise—some theory and simulation results,'' J. Acoust. Soc. Am. **95**, 21–33 (1994); J. R. Potter and M. Chitre, ''Ambient noise imaging in warm shallow seas: Second order moment and model based imaging algorithms,'' J. Acoust. Soc. Am. **106**, 3201 (1999).
[20] Y. Lee, *Statistical Theory of Communication* (Wiley, New York, 1960).
[21] R. L. Weaver and O. I. Lobkis, ''Diffuse waves in open systems and the emergence of the Greens' function,'' J. Acoust. Soc. Am. **116**, 2731–2734 (2004).
[22] N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, ''High -resolution surface-wave tomography from ambient seismic noise,'' Science **307**, 1615–1618 (2005).
[23] K. G. Sabra, P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, ''Extracting time-domain Green's function estimates from ambient seismic noise,'' AGU Geophys. Res. Lett. **32**, L03310 (2005).
[24] P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part I, Chap. 7.
[25] Closely related to conservation of energy and to reciprocity, this is one of a set of identities based on underlying symmetries that are used to constrain perturbation theories in quantum dynamics.

Ultrasonic particle concentration in a line-driven cylindrical tube

Gregory Goddard and Gregory Kaduchak
Los Alamos National Laboratory, Electronic and Electrochemical Materials and Devices Group, Mail Stop D-429, P.O. Box 1663, Los Alamos, New Mexico 87545

(Received 15 November 2004; revised 17 March 2005; accepted 17 March 2005)

Acoustic particle manipulation has many potential uses in flow cytometry and microfluidic array applications. Currently, most ultrasonic particle positioning devices utilize a quasi-one-dimensional geometry to set up the positioning field. A transducer fit with a quarter-wave matching layer, locally drives a cavity of width one-half wavelength. Particles within the cavity experience a time-averaged drift force that transports them to a nodal position. Present research investigates an acoustic particle-positioning device where the acoustic excitation is generated by the entire structure, as opposed to a localized transducer. The lowest-order structural modes of a long cylindrical glass tube driven by a piezoceramic with a line contact are tuned, via material properties and aspect ratio, to match resonant modes of the fluid-filled cavity. The cylindrical geometry eliminates the need for accurate alignment of a transducer/reflector system, in contrast to the case of planar or confocal fields. Experiments show that the lower energy density in the cavity, brought about through excitation of the whole cylindrical tube, results in reduced cavitation, convection, and thermal gradients. The effects of excitation and material parameters on concentration quality are theoretically evaluated, using two-dimensional elastodynamic equations describing the fluid-filled cylindrical shell with a line excitation.
[DOI: 10.1121/1.1904405]

PACS numbers: 43.20.Ks, 43.25.Qp, 43.25.Gf, 43.40.Rj, 43.40.At [MFH] Pages: 3440–3447

I. INTRODUCTION

The ability to discriminate and quantify distinct populations of cells/cell organelles has become increasingly important with the growing trend to focus biological studies on various cell types.[1] Flow based cytometry and cell sorting are unique techniques that permit the identification, analysis and purification of cells based on their expression of specific markers. A flow cytometer uses hydrodynamic focusing to tightly focus a sample stream, which is probed by highly focused laser beams. The high analysis rates, lack of separation steps, and varied applications ranging from medicine to homeland defense have made flow cytometers a fixture in nearly every research and diagnostic laboratory.[2] However, the size, expense and requirement for large amounts of consumables have limited flow cytometry to formal laboratory settings.

Acoustic focusing has the capability to localize particles in a similar arrangement to hydrodynamic focusing. The use of acoustic standing waves to concentrate initially homogeneously suspended particles at acoustic pressure nodal or antinodal planes in a fluid was first described by Kundt and Lehmann.[3] The effect was initially used only in the visualization of ultrasonic waves. However, the acoustic force can be used to position,[4–11] concentrate,[12–16] or fractionate[17–26] particles noninvasively. Existing ultrasonic particle manipulation devices primarily utilize one of three geometries: confocal,[4] planar,[8–21,23–26] or slowly moving wave envelope.[7,22] Inherent to the focused and planar geometries is a requirement for accurate transducer alignment. Another approach utilizes surface waves to concentrate particles.[27]

The present research investigates an alternative acoustic particle-positioning device where the acoustic excitation is generated by the entire structure, as opposed to a localized transducer. The lowest-order structural modes of a long cylindrical glass tube driven by a piezoceramic with a line contact are tuned, via material properties and aspect ratio, to efficiently couple energy from the walls of the tube into the cavity. The cylindrical geometry eliminates the need for accurate alignment of a transducer/reflector system. Experiments show that a lower energy density in the cavity, brought about through excitation of the whole cylinder, results in reduced cavitation, convection, and thermal gradients.

In Sec. II A 1, a two-dimensional vibration model of the cylindrical glass tube with line excitation, based on elastodynamic equations, is described. Section II A 2 presents a theoretical evaluation of the effect of excitation and material parameters on concentration quality and acoustic force experienced by particles in the cavity. In Sec. II B, a comparison of theoretical predictions and experimental results is performed. Dipole vibration of the external surface of the glass tube is shown to correspond to excitation frequencies of successful particle concentration.

II. METHODS

A. Theory

Consider a two-layered cylindrical system as shown in Fig. 1. This system can be modeled using the direct global matrix approach described by Ricks and Schmidt[28] applied to a two-layered system. The layers are numbered $n=1$ and $n=2$ for the cavity and tube, respectively. The variables R_1

 0001-4966/2005/117(6)/3440/8/$22.50

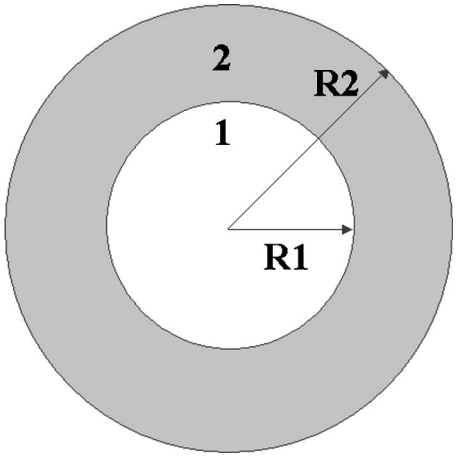


FIG. 1. Diagram of layer numbering used for modeling of cylindrical system with two layers. The fluid filled cavity is at the center contained within radius R1, the glass tube located between radii R1 and R2, and the pressure release boundary at R2.

and R_2 are the internal and external tube radii respectively. The layer numbers are indicated in the diagram.

All layers are assumed to be isotropic and homogeneous viscoelastic with Lame constants λ_n and μ_n and density ρ_n. The subscript refers to the layer number. If the layer is a solid then the displacement field $\mathbf{u}_n$ is governed by the 3D equations of elastodynamics

$$(\lambda_n+2\cdot\mu_n)\nabla\nabla\cdot\boldsymbol{u}_n-\mu_n\nabla\times\nabla\times\boldsymbol{u}_n+f_n=\rho_n\ddot{\boldsymbol{u}}_n, \tag{1}$$

where the variable f_n refers to the applied force per unit volume in layer *n*. The compressional and transverse wave speeds in layer n are related to the Lame coefficients as described

$$c_{ln}=\sqrt{(\lambda_n+2\mu_n)/\rho_n},\qquad c_{sh}=\sqrt{\mu_n/\rho_n}. \tag{2}$$

The corresponding wave numbers h_n and k_n, compressional and transverse wave numbers respectively, are expressed in terms of the angular frequency ω and the sound speeds, as shown

$$h_n=\omega/c_{ln},\qquad k_n=\omega/c_{sn}. \tag{3}$$

Following Ricks and Schmidt, in fluids, any terms involving c_{sn} and k_n are ignored and $\mu_n=0$.

A time dependent ring force of $e^{-i\omega t}$ can be assumed without loss of generality, since the time dependence of other forcing functions, including a point or line excitation, can be synthesized from time-harmonic forces by using a Fourier integral over frequency. Similarly angular dependence of $e^{i\nu\theta}$ is assumed, where ν is the order. The displacement field can be expressed as the linear superposition of homogeneous and particular solutions. The homogeneous solutions are source-free waves. The particular solutions are expressed as ring forces that would emanate from layer *n* if the layer extended over all space. The homogeneous field is governed by

$$(\lambda_n+2\cdot\mu_n)\nabla\nabla\cdot\boldsymbol{u}_n^H-\mu_n\nabla\times\nabla\times\boldsymbol{u}_n^H+\rho_n\omega^2\boldsymbol{u}_n^H=0. \tag{4}$$

The field can be expressed in terms of scalar displacement potentials Φ, Ψ, and X that satisfy the Helmholtz equations

$$(\nabla^2+h_n)\Phi_n=0,\qquad (\nabla^2+k_n)\Psi_n=0,\qquad (\nabla^2+k_n)X_n=0. \tag{5}$$

The problem can be further reduced to radial and azithmul coordinates and the homogeneous field equation can be rewritten

$$\begin{aligned}
u_{rn}&=\frac{d\Phi_n}{dr}+\frac{iv}{r}\Psi_n,\\
u_{\theta n}&=\frac{iv}{r}\Phi_n-\frac{d\Psi_n}{dr},\\
\sigma_{rrn}&=-\lambda_n h_n^2\Phi_n+2\mu_n\left(\frac{d^2\Phi_n}{dr^2}+\frac{iv}{r^2}\Psi_n+\frac{iv}{r}\frac{d\Psi_n}{dr}\right),\\
\sigma_{r\theta n}&=\mu_n\left(\frac{2iv}{r}\left(-\frac{\Phi_n}{r}+\frac{d\Phi_n}{dr}\right)-\frac{v^2}{r^2}\Psi_n+\frac{1}{r}\frac{d\Psi_n}{dr}\right.\\
&\quad\left.-\frac{d^2\Psi_n}{dr^2}\right).
\end{aligned} \tag{6}$$

Similarly the particular solutions can be expressed as

$$\begin{aligned}
\boldsymbol{u}_I^P&=\nabla\Phi_I+\nabla\times(a_z\Psi_I)+\frac{1}{k}\nabla\times\nabla\times(\boldsymbol{a}_zX_I),\\
&\quad (R_2-\epsilon)\leqslant r\leqslant R_2,\\
\boldsymbol{u}_E^P&=\nabla\Phi_E+\nabla\times(a_z\Psi_E)+\frac{1}{k}\nabla\times\nabla\times(\boldsymbol{a}_zX_E),\\
&\quad R_2\leqslant r\leqslant(R_2+\epsilon).
\end{aligned} \tag{7}$$

The subscripts *E* and *I* refer to the interior and exterior regions, $\mathbf{a}_z$ is the unit vector in the *z* direction, and ϵ refers to an infinitesimal interval. To contain how the potentials fit together at $r=R_2$, we assume continuity of stress and displacement. Then we integrate each of the vector components over a vanishingly small interval of *r*.

1. Numerical solutions

Although any pair of the Bessel and Hankel functions *J*, *Y*, $H^{(1)}$, or $H^{(2)}$ would satisfy the differential equations, desire for numerical stability dictates that we represent the homogeneous and particular solutions as

$$\Phi^P=\sum_{v=0}^{\infty}\left(\frac{iR_2F_r}{4\rho_2\omega^2}\left[\frac{dH_v^{(1)}(h_2r)}{dr}\right]_{r=R_2}J_v(h_2r)\right)e^{i(v\theta-\omega t)},$$

$$\Psi^P=\sum_{v=0}^{\infty}\left(\frac{vR_2F_r}{4\rho_2\omega^2}H_v^{(1)}(k_2R_2)J_v(k_2r)\right)e^{i(v\theta-\omega t)},$$

$$\Phi_n^H = \sum_{v=0}^{\infty} \left(A_{nv} \frac{H_v^{(1)}(h_{rn} r)}{H_v^{(1)}(h_{rn} R_{n-1})} + B_{nv} J_v(h_{rn} r) \times H_v^{(1)}(h_{rn} R_n) \right) e^{i(v\theta - \omega t)},$$

$$\Psi_n^H = \sum_{v=0}^{\infty} \left(C_{nv} \frac{H_v^{(1)}(k_{rn} r)}{H_v^{(1)}(k_{rn} R_{n-1})} + D_{nv} J_v(k_{rn} r) \times H_v^{(1)}(k_{rn} R_{n-1}) \right) e^{i(v\theta - \omega t)}. \quad (8)$$

This formulation of the solutions, defined by Ricks and Schmidt[28] in their paper, affords numerical stability over a wide range of frequencies and circumferential orders. This is achieved in three requirements on the global matrix definition.

The first is a requirement that evanescence across layers should decouple the global system into separate subsystems, so solution values caused by sources on one side of the layer should have vanishing influence on sources on the opposite side of the evanescent layer.

The second requirement on the global matrix is the choice of the pair of cylindrical Bessel and Hankel functions used in the expression of the solutions for the purpose of achieving numerical stability.

The third requirement is for the selection of normalization terms that cancel overflow and underflow conditions. Normalization terms consisting of Hankel functions of the first kind evaluated at inner or outer radii of each layer were alternately multiplied and divided by each of the terms.

The coefficients $A_{n\nu}$, $B_{n\nu}$, $C_{n\nu}$, and $D_{n\nu}$ will be determined using the boundary conditions for each layer. The coefficient F_r is the force amplitude. Several assumptions serve to simplify the problem and reduce the size of the global matrix. Since the Hankel function is undefined near $r=0$, $A_{1\nu}=0$. Since inviscid fluids do not support transverse waves, $C_{1\nu}=0$ and $D_{1\nu}=0$. These conditions are true for all values of ν.

The global matrix is constructed using the unknown displacements and stresses associated with the homogeneous waves, written in terms of an amplitude vector and set equal to the displacements and stresses due to the particular solutions derived from the forcing excitation. Solutions for the coefficients are determined performing Gaussian elimination with partial pivoting or equivalently by applying Cramer's rule to the global matrix equation,

$$\begin{bmatrix} g_{11} & g_{12} & g_{13} & g_{14} & g_{15} \\ g_{21} & g_{22} & g_{23} & g_{24} & g_{25} \\ g_{31} & g_{32} & g_{33} & g_{34} & g_{35} \\ g_{41} & g_{42} & g_{43} & g_{44} & g_{45} \\ g_{51} & g_{52} & g_{53} & g_{54} & g_{55} \end{bmatrix} \begin{bmatrix} B_1 \\ A_2 \\ B_2 \\ C_2 \\ D_2 \end{bmatrix}_v = [\text{bnd. cond.}]. \quad (9)$$

Both techniques work reliably if the equations are scaled to make all of the matrix coefficients dimensionless. To do so, we divide the displacement equations by the compression wave number of the fluid layer, h_1 and divide all stress equations by $\lambda_1 h_1^2$. The choice of the reference layer is not critical, but the fluid layer is chosen since the Lame' coefficient for a fluid only contains compressional sound speed, and therefore simplifies the equation.

The boundary conditions, which generate the global matrix, are

$$\begin{aligned} u_{r1}^H(R_1) &= u_{r2}^H(R_1), \\ \sigma_{r1}^H(R_1) &= \sigma_{r2}^H(R_1), \\ \sigma_{r\theta 2}^H(R_1) &= 0, \\ \sigma_{rr2}^H(R_2) &= \sigma_{rr}^P(R_2), \\ \sigma_{r\theta 2}^H(R_2) &= \sigma_{r\theta}^P(R_2), \end{aligned} \quad (10)$$

The particular solutions given assume a delta function excitation point source. To account for the finite width of the element used in the experiments, a Gaussian weighting function was applied to the forcing function. For aesthetic purposes, the region about $\theta = 3\pi/2$ radians was chosen as the excitation point.

2. Acoustic force on particles

Particles suspended in the fluid-filled cylindrical cavity and exposed to an acoustic standing wave field experience a time-averaged drift force that transports them to a nodal or antinodal position depending on the relationships between the density and acoustic velocity of the particle and fluid, also called the acoustic contrast ratio.[29–33] For the purposes of this research, the relative acoustic properties of the particles of interest, namely bacteria, spores and other biological cells, to the suspending fluid, such as buffered saline solutions, yield acoustic contrast ratios that drive the particles to the pressure node.

Forces also are present in the sound field that exert torque on particles, inducing spin or alignment. Second order forces between particles[34] such as Konig or Bjerknes force can additionally serve to aggregate the particles into concentrated clumps. Only the primary acoustic force will be considered here.

King[29] performed the first quantitative analysis of the primary radiation force on a particle in a sound field. He considered the radiation force only, assuming incompressible spheres, noting that the radiation force on particles with radii less than a wavelength is greater in a standing than a traveling wave field. Yosioka *et al.*[30,32,33] and later Gorkov[31] extended the analysis to include the influence of particle compressibility on the force moving the particles to nodal or antinodal positions. For a dilute suspension in an arbitrary field, Gorkov's theory for noninteracting particles provides a good description of the equilibrium particle distribution. The acoustic force **F** acting on a small spherical particle of radius r and density ρ_p in a fluid of density ρ_f is given by

$$\mathbf{F} = -\nabla \left(\frac{2}{3} \pi R^3 \left[\frac{\overline{p^2}}{\rho_f c_f^2} - \frac{\overline{p^2}}{\rho_p c_p^2} - \frac{3\rho_f(\rho_p - \rho_f)}{2\rho_p + \rho_f} \overline{v^2} \right] \right), \quad (11)$$

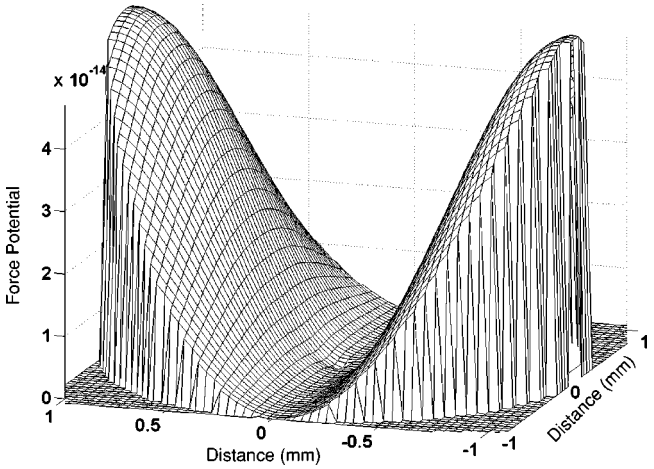


FIG. 2. Plot of predicted acoustic potential experienced by polystyrene microspheres in water. The water is excited by the coupled tube/water dipolar mode. The vertical axis is labeled in arbitrary units of potential energy. The horizontal axes denote distance from the center of the tube in millimeters. The *x*- and *y*-axes are at the right and bottom of the figure respectively. The source excitation is located at $(0,-1.985,0)$.

where the variables c_f and c_p are the acoustic velocities in the fluid and the particle, respectively, p and v are the pressure and velocity of the wave at the location of the particle.

Combining the expression in Eq. (11) with the model predictions allowed for determination of vibrational modes of the system that would efficiently concentrate particles to the central axis. The coupled tube/fluid dipolar mode is found to most efficiently concentrate particles on the axis. Figure 2 presents the potential experienced by the particles in the cavity. Force is equal to the negative gradient of the potential.

B. Experiment

A type 4A lead zirconate titanate (PZT) piezoceramic crystal is glued axially to a piece of soda lime glass tube with an inner diameter of 2.2 mm, and an outer diameter of 3.97 mm. The crystal is 30 mm long, 3 mm thick, and 1.5 mm wide and has a thickness mode resonance of 420 kHz. A diagram of the tube is shown in Fig. 3. Initially, the resonance of the system is determined to be approximately 417 kHz by scanning the drive frequency to find the frequency at which particles concentrate to the central axis of the tube. In subsequent experiments, maximization of a received signal on another similar piece of piezoceramic placed inline with the driving transducer along the axis of the tube was used for tuning. The system was driven using a SRS D5345 function generator, which was then passed through a Krohnhite 7500 power amplifier. The received signal was viewed on an oscilloscope.

A section of thick-walled quartz tube is compared with

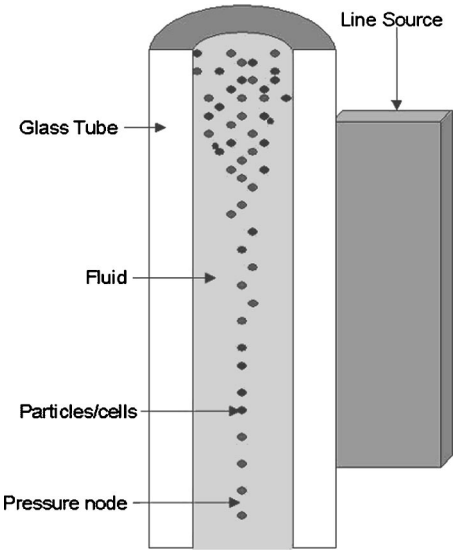


FIG. 3. Diagram of tube used in experiments indicating the locations of the piezoceramic line source, fluid, particles, and pressure node. The predicted and experimentally verified spatial location of concentration for the particles at the pressure node is also indicated.

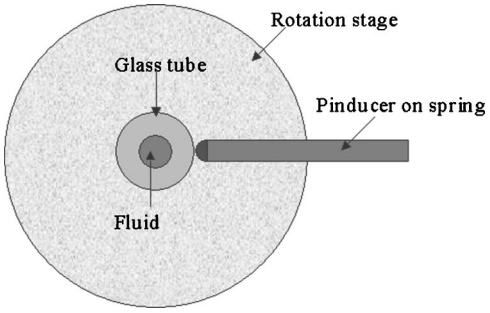


FIG. 4. Diagram of angle scan apparatus used to measure surface vibration of the glass tube under ultrasonic excitation for verification of model predictions. A pinducer is mounted to a spring to assure constant contact with the tube. The tube is mounted at the center of the rotation stage. The tip of the pinducer is attached to a metal hemisphere to assure minimal angular integration as the tube is rotated about its central axis.

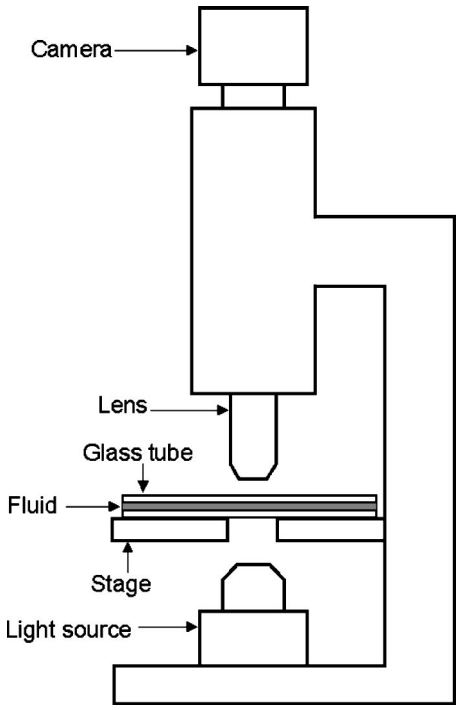


FIG. 5. Diagram of bottom-illuminated microscope used to image concentration pattern in a tube. A high resolution black and white CCD camera (indicated in figure) is used to capture the images via a computer. Particles are imaged with and without fluid flow through the tube. The tube is mounted to the microscope stage.

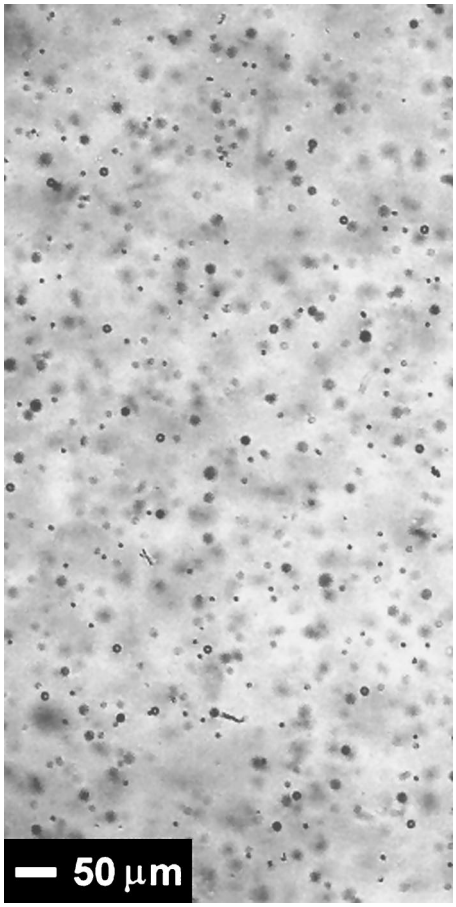


FIG. 6. Micrograph of unconcentrated 10-micron polystyrene particles suspended in water. The particle suspension is in the soft glass tube.

the soft glass tube. The quartz tube has an outer diameter of 7.85 mm and an inner diameter of 2.0 mm. An identical piezoceramic is used as a line source on the quartz tube in the same configuration as soft glass.

Typical drive signal amplitudes into the piezoceramic are $10-12\,V_{pp}$ and 80 mA for both the soft glass and quartz tubes. The current is measured as a voltage across a 10-ohm resistor in series with the piezoceramic. Ten-micron polystyrene particles diluted to a concentration of approximately 0.025% by volume in water are flowed through the tube using a gravity feed system. Flow rates of 5–25 mm/s are used. The liquid was only minimally stirred to maintain suspension of the particles in a feed reservoir. A peristaltic pump is used to refill the feed reservoir from the sink reservoir.

For the determination of tube surface vibration, the tube is mounted to a computer controlled angular stepper motor stage and probed with a pinducer. A diagram of the angle scanning apparatus is shown in Fig. 4. A half-hemisphere of solder is affixed to the end of the pinducer to assure point contact with the tube, and thus minimize angular integration of the signal. To maintain constant contact with the tube, the pinducer was mounted on a spring. A computer controlled HP Infiniuum 54810A oscilloscope is used for data collection. The tube is characterized empty and filled with water. Images of the concentration pattern are taken using a bottom-illuminated microscope and 1280×1024 12-bit digital camera, diagrammed in Fig. 5.

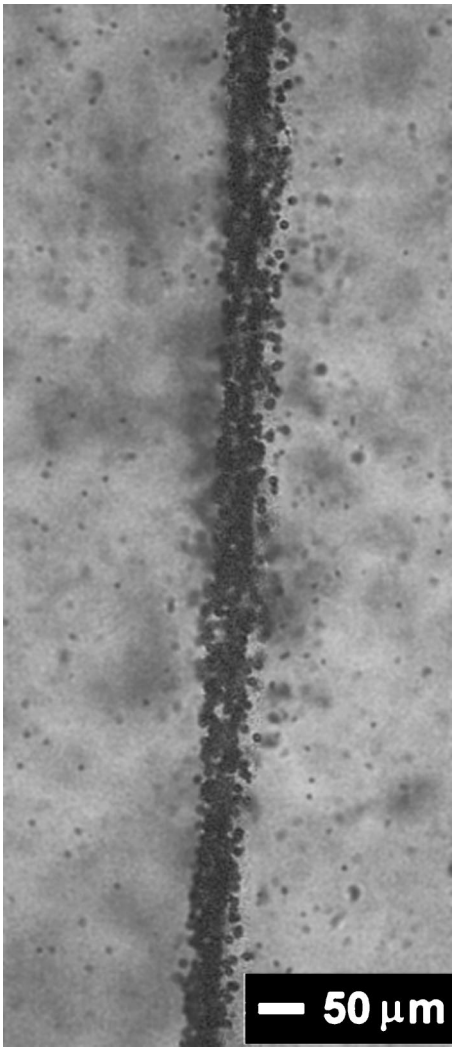


FIG. 7. Micrograph of acoustically concentrated 10-micron particles in a soft glass tube that is excited at the predicted and experimentally verified tube/fluid dipolar mode for the soft glass tube at frequency 417 kHz.

III. RESULTS AND DISCUSSION

The tube is driven at 0.8–0.9 W to concentrate over a 15 cm length with a single drive transducer. A typical concentration pattern is shown in Fig. 7. Particles are ten microns in diameter with a standard deviation of 0.7 microns. The suspending fluid is distilled water. No degassing is performed of the liquid in order to most accurately mimic the conditions expected in a microfluidic system. The concentration pattern took approximately 5 s to form from an initially homogeneous suspension similar to that shown in Fig. 6. Larger particles visible in the micrograph are agglomerations of particles.

As can be seen in Fig. 7, the concentrating line is only a few particle diameters across. Particles visible on either side of the concentrated line are adhered to the tube wall by electrostatic charge. Various concentrations of beads ranging from 0.02% to 0.2% by weight are investigated. No significant differences are observed in concentration times, but higher concentrations of particles also showed increased agglomeration. The agglomerations are also concentrated at the center, but did not disassociate at power levels less than 1 W.

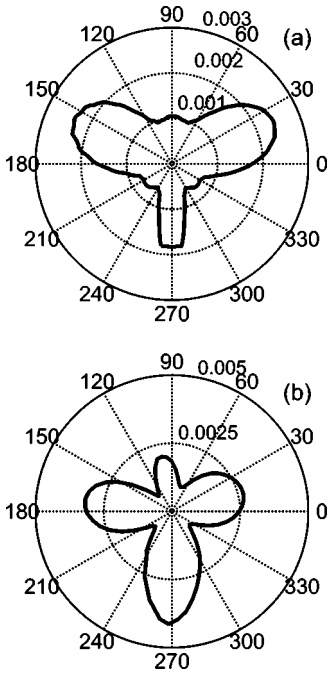


FIG. 8. Polar plot of surface displacement for empty soft glass tube excited at 417 kHz. Radial coordinates in arbitrary units of amplitude, angular coordinates in degrees. (A) Predicted; (B) Experimentally measured.

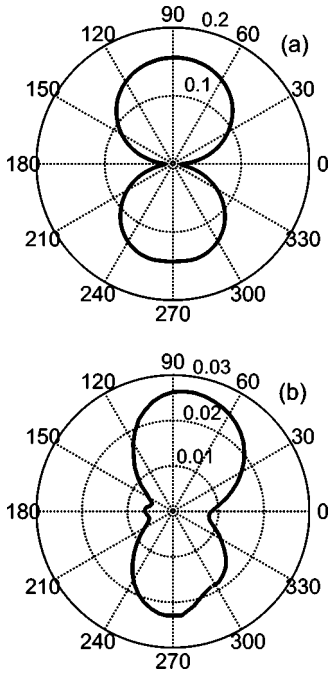


FIG. 9. Polar plot of surface displacement for a soft glass tube filled with water and excited at 417 kHz. Radial coordinates in arbitrary units of amplitude, angular coordinates in degrees. (A) Predicted; (B) Experimentally measured.

The outer surface displacement of the soft glass tube is calculated at 417 kHz. The material properties of the soft glass are determined by matching index of refraction, density, compressional sound speed, and coefficient of thermal expansion to soda lime glass. A compressional sound speed of 5900 m/s, transverse sound speed of 3300 m/s, and density of 2.43 g/cm^3 are used for the soft glass. The values of sound speed and density of air used in the calculations are 340 m/s and 10^{-6} g/cm^3, respectively. Water is assumed to have sound speed of 1487 m/s and density of 1 g/cm^3.

Predicted surface displacement for empty soft glass tube is shown in Fig. 8(A). The corresponding measured surface displacement is shown in Fig. 8(B). In Fig. 8(A), the radial axis demarcations correspond to predicted surface vibration for unitary excitation amplitude. The radial demarcations in Fig. 8(B) are given in received voltage from the probing pinducer. The demarcations around the polar axis are given in degrees in both figures.

Three primary lobes at approximately 120° relative angles and a directly opposite smaller lobe can be seen in calculated and measured results as evidenced in Fig. 8. The greater angular spread in the measured data is due to width of angular coupling of the source. When the tube is filled with water, the four lobed external displacement collapsed to a strong dipole, as can be seen in the calculated results in Fig. 9(A). The experimental displacement measurement for the water filled tube also yielded a dipole pattern, as can be seen in Fig. 9(B).

It should be noted that the predicted vibration amplitude of the tube when empty versus filled with water is approximately two orders of magnitude larger for the water-filled case for equal drive amplitude. The measured vibration amplitude for the two cases also shows significant increase in amplitude when the cavity is filled with water. This would suggest a greater energy coupling into the structure when water is present in the cavity. A thick-walled quartz tube was obtained for comparison with the soft glass tube. The material properties of quartz used are as follows: compressional sound speed of 5700 m/s, transverse sound speed of 2650 m/s, and density of 2.65 g/cm^3. The same properties of air and water as used in the soft glass calculations are assumed for quartz tube. The model predicted a dipole at 462 kHz for the quartz tube, as can be seen in Fig. 10(A). Surface vibration measurement yielded the results shown in Fig. 10(B). As

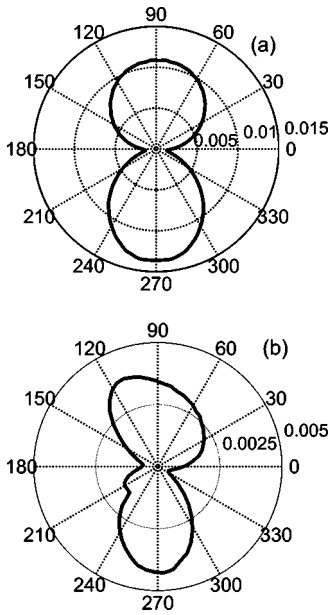


FIG. 10. Polar plot of surface displacement for quartz tube filled with water and excited at 462 kHz. Radial coordinates in arbitrary units of amplitude, angular coordinates in degrees. (A) Predicted; (B) experimentally measured.

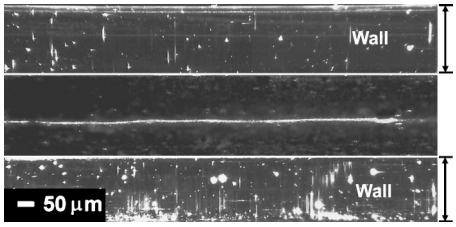


FIG. 11. Micrograph of acoustically concentrated 10-micron particles in a quartz tube that is excited at the predicted and experimentally verified tube/fluid dipolar mode for the thick quartz tube at a frequency of 462 kHz. Small asymmetries in the tube caused a perturbation to the acoustic potential which consequently displaced the local potential minima for the particles off of the central axis of the tube.

can be seen, there is good correspondence between the predicted and experimental results. Particles were fed through the quartz tube and the concentration pattern at 462 kHz is imaged using the bottom-illuminated microscope. A micrograph of the concentration pattern is shown in Fig. 11. The 2 mm thick walls can be seen at the top and bottom of the micrograph. As can be seen in the micrograph, the dipole does indicate a frequency at which concentration takes place. However, small asymmetries in the tube caused a perturbation to the acoustic potential which consequently displaced the local potential minima for the particles away from the central axis of the tube.

Acoustic cavitation has been shown to affect the viability of bacterial cells similar to those of interest in this research.[35–37] Although the sample fluid was not degassed, which would be expected to raise the incidence of acoustic cavitation, particularly within the frequency range used in these experiments,[38–40] no cavitation was observed during the experiments. By exciting the whole cylindrical tube, a lower energy density within the fluid is achieved. This allows for particle concentration without exceeding the acoustic cavitation threshold.

IV. CONCLUSION

The most common premise for low power acoustic concentration is the requirement of quarter wave matching layers. The inherent alignment problems of either planar or confocal geometries, even in the traveling wave case, for proper positioning, make the process more difficult. The possibility of strongly exciting cavity modes by driving at far below quarter wave thickness of a cylinder wall, offers some intriguing possibilities for applications. The external line drive and use of the lowest order modes of a cylinder allow for application to continuous flow-through systems without necessity of a sedimentation step or second drive transducer for traveling wave creation. Concentration times are approximately 5–10 s at 25 mm/s. Correspondence of concentration to a dipole mode of the system is shown. Reduction of the elastodynamic equations to two dimensions has proven to be a reasonable simplification. The validity of the global matrix model for the vibration has also been demonstrated. Application can be made of this model to determining optimal material properties and geometric parameters for particle concentration. Although power consumption for the set experiments described herein was approximately 1 W, concentration is seen at lower power levels.

ACKNOWLEDGMENTS

We are grateful to Dave Lizon and Roger Lujan for help with the mechanical design, and Dr. Dipen Sinha and Dr. Christopher Kwiatkowski for many helpful discussions. This research was funded by LANL Laboratory Directed Research and Development funds.

[1] H. M. Shapiro, *Practical Flow Cytometry* (Wiley, Hoboken, 2003).
[2] S. Graves, J. Nolan, J. Jett, J. Martin, and L. Sklar, ''Nozzle design parameters and their effects on rapid sample delivery in flow cytometry,'' Cytometry **47**, 127–137 (2002).
[3] A. Kundt and O. Lehmann, ''Longitudinal vibrations and acoustic figures in cylindrical columns of liquids,'' Ann. Phys. **153**, 1–11 (1874).
[4] H. M. Hertz, ''Standing-wave acoustic trap for nonintrusive positioning of microparticles,'' J. Appl. Phys. **78**, 4845–4849 (1995).
[5] A. Haake and J. Dual, ''Micro-manipulation of small particles by node position control of an ultrasonic standing wave,'' Ultrasonics **40**, 317–322 (2002).
[6] M. Takeuchi and K. Yamanouchi, ''Ultrasonic micromanipulation of small particles in liquid,'' Jpn. J. Appl. Phys., Part 1 **33**, 3045–3047 (1994).
[7] Y. Yamakoshi and Y. Noguchi, ''Micro particle trapping by opposite phases ultrasonic traveling waves,'' Ultrasonics **36**, 873–878 (1998).
[8] W. T. Coakley, D. W. Bardsley, M. A. Grundy, F. Zamani, and D. J. Clarke, ''Cell manipulation in ultrasonic standing wave fields,'' J. Chem. Technol. Biotechnol. **44**, 43–62 (1989).
[9] W. T. Coakley, G. Whitworth, M. A. Grundy, R. K. Gould, and R. Allman, ''Ultrasonic manipulation of particles and cells,'' Bioseparation **4**, 73–78 (1994).
[10] J. J. Hawkes, D. Barrow, and W. T. Coakley, ''Microparticle manipulation in millimetre scale ultrasonic standing wave chambers,'' Ultrasonics **36**, 925–931 (1998).
[11] W. T. Coakley, J. J. Hawkes, M. A. Sobanski, C. M. Cousins, and J. Spengler, ''Analytical scale ultrasonic standing wave manipulation of cells and microparticles,'' Ultrasonics **38**, 638–641 (2000).
[12] K. Yasuda, S. S. Haupt, and S. Unemura, ''Using acoustic radiation force as a concentration method for erythrocytes,'' J. Acoust. Soc. Am. **102**, 642–645 (1997).
[13] K. Yasuda, ''Nondestructive, noncontact handling and concentration of μm-sized particles in liquid using superposition method of ultrasound,'' in *Micro-Electro-Mechanical Systems (MEMS)-2000*, ASME International Mechanical Engineering Congress and Exposition, November 2000, Orlando (2000), pp. 393–396.
[14] P. Vainshtein, M. Fichman, K. Shuster, and C. Gutfinger, ''The effect of centerline particle concentration in a wave tube,'' J. Fluid Mech. **306**, 31–42 (1996).
[15] K. Yasuda, S. Unemura, and K. Takeda, ''Concentration and fractionation of small particles in liquid by ultrasound,'' Jpn. J. Appl. Phys., Part 1 **34**, 2715–2720 (1995).
[16] Goldshtein, K. Shuster, P. Vainshtein, M. Fichman, and C. Gutfinger, ''Particle motion in resonance tubes,'' J. Fluid Mech. **360**, 1–20 (1998).
[17] T. L. Tolt and D. L. Feke, ''Separation devices based on forced coincident response of fluid-filled pipes,'' J. Acoust. Soc. Am. **91**, 3152–3156 (1992).
[18] T. L. Tolt and D. L. Feke, ''Separation of dispersed phases from liquids in acoustically driven chambers,'' Chem. Eng. Sci. **48**, 527–540 (1993).
[19] S. Gupta and D. L. Feke, ''Fractionation of mixed particulate solids according to compressibility using ultrasonic standing wave fields,'' Chem. Eng. Sci. **50**, 3275–3284 (1995).
[20] M. C. Bekker, J. P. Meyer, L. Pretorius, and D. F. Van Der Merwe, ''Separation of solid-liquid suspensions with ultrasonic acoustic energy,'' Water Res. **31**, 2543–2549 (1997).
[21] G. Whitworth, M. A. Grundy, and W. T. Coakley, ''Transport and harvesting of suspended particles using modulated ultrasound,'' Ultrasonics **29**, 439–444 (1991).
[22] G. Whitworth and W. T. Coakley, ''Particle column formation in a stationary ultrasonic field,'' J. Acoust. Soc. Am. **91**, 79–85 (1992).
[23] J. J. Hawkes, D. Barrow, J. Cefai, and W. T. Coakley, ''A laminar flow expansion chamber facilitating downstream manipulation of particles con-

centrated using an ultrasonic standing wave,'' Ultrasonics **36**, 901–903 (1998).
[24] T. Masudo and T. Okada, ''Particle Characterization and Separation by a Coupled Acoustic—Gravity Field,'' Anal. Chem. **73**, 3467–3471 (2001).
[25] R. K. Gould and W. T. Coakley, ''The effects of acoustic forces on small particles in suspension,'' in *Proceedings of the 1973 Symposium on Finite Amplitude Wave Effects in Fluids*, edited by L. Bjorno (Pergamon, Guildford, 1974), pp. 252–257.
[26] K. Higashitani, M. Fukushima, and Y. Matsuno, ''Migration of suspended particles in plane stationary ultrasonic field,'' Chem. Eng. Sci. **36**, 1187–1192 (1981).
[27] A. Haake and J. Dual, ''Positioning of small particles by an ultrasound field excited by surface waves,'' Ultrasonics **42**, 75–80 (2004).
[28] D. C. Ricks and H. Schmidt, ''A numerically stable global matrix method for cylindrically layered shells excited by ring forces,'' J. Acoust. Soc. Am. **95**, 3339–3349 (1994).
[29] L. V. King, ''On the acoustic radiation on spheres,'' Proc. R. Soc. London, Ser. A **147**, 212–240 (1933).
[30] K. Yosioka and Y. Kaisima, ''Acoustic radiation pressure on a compressible sphere,'' Acustica **5**, 167–173 (1955).
[31] L. P. Gor'kov, ''On the forces acting on a small particle in an acoustical field in and ideal fluid,'' Sov. Phys. Dokl. **6**, 773–775 (1962).
[32] T. Hasegawa and K. Yosioka, ''Acoustic radiation force on a solid elastic sphere,'' J. Acoust. Soc. Am. **46**, 1139–1143 (1969).
[33] K. Hasegawa and K. Yosioka, ''Acoustic radiation force on fused silica spheres, and intensity determination,'' J. Acoust. Soc. Am. **58**, 581–585 (1975).
[34] A. Doinikov, ''Acoustic radiation interparticle forces in a compressible fluid,'' J. Fluid Mech. **444**, 1–21 (2001).
[35] E. Joyce, S. Phull, J. Lorimer, and T. Mason, ''The development and evaluation of ultrasound for the treatment of bacterial suspensions: A study of frequency, power and sonication time on cultured bacillus species,'' Ultrason. Sonochem. **10**(6), 315–318 (2003).
[36] A. Vollmer, S. Kwakye, M. Halpern, and E. Everbach, ''Bacterial stress responses to 1-Megahertz pulsed ultrasound in the presence of microbubbles,'' Appl. Environ. Microbiol. **64**, 3927–3931 (1998).
[37] G. Scherba, R. Weigel, and W. O'Brien, ''Quantitative assessment of the germicidal efficacy of ultrasonic energy,'' Appl. Environ. Microbiol. **57**, 2079–2084 (1991).
[38] R. Apfel, ''Acoustic cavitation prediction,'' J. Acoust. Soc. Am. **69**, 1624–1633 (1981).
[39] H. Flynn, ''Physics of acoustic cavitation in liquids,'' *Physical Acoustics* (Academic, New York, 1964), Vol. 1B, pp. 57–172.
[40] E. Neppiras, ''Acoustic Cavitation,'' Phys. Rep. **61**, 160–251 (1980).

Vortical and acoustical mode coupling inside a porous tube with uniform wall suction

T. A. Jankowski[a] and J. Majdalani[b]
University of Tennessee Space Institute, Tullahoma, Tennessee 37388

(Received 30 January 2002; revised 18 February 2005; accepted 22 March 2005)

This paper considers the oscillatory motion of gases inside a long porous tube of the closed-open type. In particular, the focus is placed on describing an analytical solution for the internal acoustico-vortical coupling that arises in the presence of appreciable wall suction. This unsteady field is driven by longitudinal oscillatory waves that are triggered by small unavoidable fluctuations in the wall suction speed. Under the assumption of small amplitude oscillations, the time-dependent governing equations are linearized through a regular perturbation of the dependent variables. Further application of the Helmholtz vector decomposition theorem enables us to discriminate between acoustical and vortical equations. After solving the wave equation for the acoustical contribution, the boundary-driven vortical field is considered. The method of matched-asymptotic expansions is then used to obtain a closed-form solution for the unsteady momentum equation developing from flow decomposition. An exact series expansion is also derived and shown to coincide with the numerical solution for the problem. The numerically verified end results suggest that the asymptotic scheme is capable of providing a sufficiently accurate solution. This is due to the error associated with the matched-asymptotic expansion being smaller than the error introduced in the Navier-Stokes linearization. A basis for comparison is established by examining the evolution of the oscillatory field in both space and time. The corresponding boundary-layer behavior is also characterized over a range of oscillation frequencies and wall suction velocities. In general, the current solution is found to exhibit features that are consistent with the laminar theory of periodic flows. By comparison to the Sexl profile in nonporous tubes, the critically damped solution obtained here exhibits a slightly smaller overshoot and depth of penetration. These features may be attributed to the suction effect that tends to attract the shear layers closer the wall. [DOI: 10.1121/1.1905639]

PACS numbers: 43.20.Mv, 43.28.Py [LLT] Pages: 3448–3458

I. INTRODUCTION

In a previous article,[1] an asymptotic solution was presented for the acoustico-vortical field that was triggered by small fluctuations in wall injection inside a porous enclosure. This effort was supplemented by a higher-order analytical approximation based on a Liouville-Green transformation.[2] The current article extends the former studies by considering the suction-driven flow analog. In particular, the focus is presently shifted to the acoustico-vortical field inside a porous tube with appreciable wall suction. Despite the apparent resemblance with the injection-driven problem, the presence of wall suction leads to a dissimilar physical setting requiring a separate mathematical treatment. The new treatment applies to the linearized Navier-Stokes equations obtained through flow decomposition.

Suction-induced flows that are susceptible to acoustical oscillations are encountered in diverse applications. One pertains to membrane filtration[3,4] and the separation of uranium isotopes by differential gas diffusion.[5–7] Another arises in the modeling of mechanically assisted respiration,[8] hemodialysis in artificial kidneys,[9] and mass transport in the lungs.[10,11] A third concerns boundary-layer separation and control.[12,13] While past investigations have primarily concentrated on nonoscillatory behavior, the intent of this article is to account for possible flow periodicity that can be often introduced either internally, through a self-sustaining mechanism, or externally, through an oscillating boundary.

The inception of unavoidable fluctuations in injection-driven flows has been reported in several experimental investigations.[14–16] One source of oscillatory behavior can be ascribed to random fluctuations in the wall injection rate. These are often inevitable and take place at random frequencies. Clearly, those matching the tube's natural frequency give rise to a self-sustaining acoustical field.[14–16] Theoretical investigations suggest that similar behavior can be expected in suction-driven flows.[17–20] In some applications, however, an externally induced oscillatory field can exist. This is true when the field is created from laboratory-controlled wave generators, or from periodic sources supplied by autonomous mechanisms. Examples of the latter include the peristaltic mechanisms in control of respiratory and circulatory functions.

Whether the time dependence comes through self-induced or externally triggered sources, the goal of this study is to incorporate these oscillations into the suction-driven

[a]Current address: Los Alamos National Laboratory, Engineering Sciences and Applications Division-Applied Engineering Technologies (ESA-AET), Los Alamos, NM 87545.
[b]Author to whom all correspondence should be addressed: Department of Mechanical, Aerospace and Biomedical Engineering, University of Tennessee (UTSI), Tullahoma, TN 37388. Electronic mail: maji@utsi.edu

flow. By virtue of the complexity of the resulting model, the outcome we seek is an approximate solution only. Unlike the mean flow counterpart,[21–24] the existence of an exact unsteady solution can be proven impossible in light of recent mathematical reports.[17–20] To the author's knowledge, the current effort may constitute the first attempt at formulating a time-dependent solution for the suction-driven field inside a porous tube.

It may be safely stated that Berman's landmark paper[5] has provided the original motivation for the suction-driven problems. This, Berman accomplished, by showing that the Navier-Stokes equations could be reduced to a nonlinear ordinary differential equation (ODE) of fourth order. Berman was also first in applying asymptotics to obtain a solution in the presence of small wall suction. In the process, Berman's perturbation parameter was quantified by the cross-flow Reynolds number $\mathrm{Re}\equiv V_s a/\nu_0$. This term was based on the wall suction velocity V_s, viscosity ν_0, and tube radius a. Based on Berman's equation, numerous studies ensued, some with the purpose of generating closed-form descriptions for suction-driven flows over different ranges of Re. By way of example, one may cite Yuan,[25,26] Sellars,[27] and Terrill,[28–30] who have extended Berman's small suction case to encompass higher ranges of Re. While Yuan extended Berman's solution up to $\mathrm{Re}=20$, Sellars and Terrill developed approximate solutions that coincided as $\mathrm{Re}\to\infty$ in both channels and tubes. In this work, their joint solution will be used to prescribe the mean flow character.

A number of mathematical studies have also addressed the important aspect of solution multiplicity.[17–24] In summary, a total of four solution types were detected in the suction-driven porous tube.[20] For small suction, two solutions were shown to exist for $0\leqslant \mathrm{Re}\leqslant 2.3$. While no solutions were present in the interval $2.3<\mathrm{Re}<9.1$, four outcomes were observed for $9.1\leqslant \mathrm{Re}<\infty$. Stability of the porous tube solution was also examined by Zaturska, Drazin, and Banks.[20] Their results have indicated that the Berman-type similarity solutions could only be stable to time-dependent perturbations in the ranges $0\leqslant \mathrm{Re}\leqslant 2.3$ and $9.113\leqslant \mathrm{Re}<44.709$. Outside these ranges, unstable solutions due to asymmetric perturbations were likely to exist. Nonetheless, despite the speculations made regarding the physicality of unstable solutions, no laboratory confirmation could be found in the technical literature. For this reason, the current study utilizes the leading-order solution for large Re as a model for the mean flow field in the presence of symmetric oscillations.

The physical idealization starts in Sec. II with a definition of the basic flow model. This is initiated with a description of the system geometry and physical criteria. In Sec. III, the governing equations are presented in their general dimensional form. Subsequently, equations and variables are normalized, linearized, and decomposed into steady and time-dependent sets.[1] The temporal field is further subdivided in Sec. IV using the Helmholtz vector superposition theorem. At the outset, acoustical and vortical equations are distinguished. While the pressure-driven response is dealt with immediately, the momentum equation obtained from the rotational component is simplified to a second-order ODE. In Sec. V the rotational momentum equation is examined before a solution is attempted. This is followed in Sec. VI by the presentation of a composite solution based on matched-asymptotic expansions. An exact series expansion is developed in Sec. VII. Finally, results are displayed and discussed in Sec. VIII. Everywhere, the scope is restricted to $\mathrm{Re}\geqslant 20$.

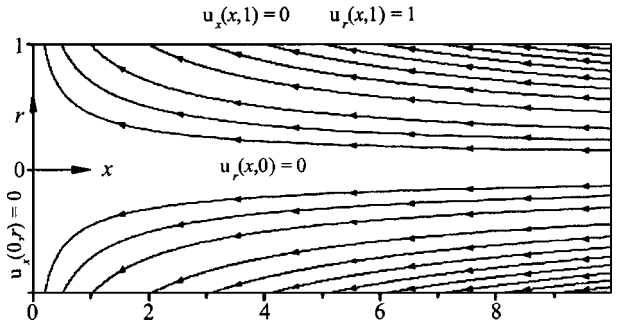


FIG. 1. Geometrical sketch showing mean flow streamlines and boundary conditions based on Terrill's large suction-flow solution.

II. THE BASIC FLOW MODEL

A. The porous tube

We shall consider a long porous tube of radius a. Fluid is extracted from the circumferential walls at a uniform wall velocity V_s. The length of the tube is defined by L and the system is simplified by imposing the condition of symmetry about the tube's axis. This choice permits the solution domain to become reducible to $0\leqslant x\leqslant l$, and $0\leqslant r\leqslant 1$, where $l=L/a$ is the dimensionless tube length. For illustrative purposes, Fig. 1 sketches a cross section of the tube with the mean flow streamlines calculated from Terrill's solution for large suction.[28]

Under the influence of small variations in the suction rate, a tube that is closed at the head end and open at the aft end can develop longitudinal pressure oscillations of amplitude A_p. The corresponding acoustical frequency can be specified by[31,32]

$$\omega_0=(m-\tfrac{1}{2})\pi a_0/L, \tag{1}$$

where a_0 is the speed of sound and m is the oscillation mode number.

B. Limiting conditions

In order to simplify the analysis to the point where an analytical solution can be arrived at, several restrictions must be observed. First, the mean flow must be Newtonian, laminar, and unsusceptible to mixing, swirling, or turbulence. Furthermore, the oscillatory pressure amplitude is taken to be small in comparison with the stagnation pressure.

III. GOVERNING EQUATIONS

A. The conservation equations

Employing asterisks to designate dimensional variables, density, pressure, time, velocity, and the shear stress tensor

can be represented by ρ^*, p^*, t^*, $\mathbf{u}^*$, and τ^*, respectively. Continuity and conservation of momentum can then be written in their general forms[33]

$$\partial\rho^*/\partial t^* + \nabla^* \cdot (\rho^* \mathbf{u}^*) = 0, \tag{2}$$

$$\partial(\rho^* \mathbf{u}^*)/\partial t^* + \nabla^* \cdot (\rho^* \mathbf{u}^* \mathbf{u}^*) = -\nabla^* p^* - (\nabla^* \cdot \tau^*). \tag{3}$$

By using continuity to simplify Eq. (3) and viscous transfer for a Newtonian fluid, one can transform Eq. (3) into

$$\rho^*[\partial \mathbf{u}^*/\partial t^* + (\mathbf{u}^* \cdot \nabla^*)\mathbf{u}^*] = -\nabla^* p^* + \mu^*[4\nabla^*(\nabla^* \cdot \mathbf{u}^*)/3 - \nabla^* \times (\nabla^* \times \mathbf{u}^*)], \tag{4}$$

where μ^* is the dynamic viscosity.

To be general, dimensionless parameters are introduced. Spatial coordinates are hence normalized by a, while velocity and time are made dimensionless by a_0 and ω_0, respectively. In summary, we let

$$x = x^*/a, \quad r = r^*/a, \quad t = \omega_0 t^*, \quad \mathbf{u} = \mathbf{u}^*/a_0,$$
$$p = p^*/(\gamma p_0), \quad \text{and} \quad \rho = \rho^*/\rho_0, \tag{5}$$

where γ is the ratio of specific heats, and ρ_0 and p_0 are the stagnation density and pressure. Following this nomenclature, Eqs. (2) and (4) become

$$\omega \partial\rho/\partial t + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{6}$$

$$\rho[\omega \partial \mathbf{u}/\partial t + (\mathbf{u} \cdot \nabla)\mathbf{u}] = -\nabla p + \mathrm{M}\varepsilon[4\nabla(\nabla \cdot \mathbf{u})/3 - \nabla \times (\nabla \times \mathbf{u})]. \tag{7}$$

Equations (6) and (7) employ the definitions of the nondimensional frequency $\omega \equiv \omega_0 a/a_0$, the suction Mach number $\mathrm{M} \equiv V_s/a_0$, and the small parameter $\varepsilon \equiv 1/\mathrm{Re}$.

B. Perturbed variables

With the introduction of small amplitude oscillations, the instantaneous pressure can be expressed as the linear sum of time-dependent and steady components:

$$p = p^{(0)} + \varepsilon_w p^{(1)} \exp(-it), \tag{8}$$

where $i = \sqrt{-1}$, $\varepsilon_w = A_p/(\gamma p_0)$, and $[p^{(0)}, p^{(1)}]$ are spatial functions. Unlike $p^{(1)}$, $p^{(0)}$ can be shown to be a constant at order $\mathcal{O}(\mathrm{M}^2)$, namely, $p^{(0)} = 1/\gamma + \mathcal{O}(\mathrm{M}^2)$. Noting that the mean flow solution is incompressible, small compressibility effects can only influence the time-dependent field. Density can thus be normalized by its mean component and expanded in a similar fashion viz.

$$\rho = 1 + \varepsilon_w \rho^{(1)} \exp(-it). \tag{9}$$

The total velocity can also be expanded as

$$\mathbf{u} = \mathrm{M}\mathbf{U} + \varepsilon_w \mathbf{u}^{(1)} \exp(-it), \tag{10}$$

where $\mathbf{U}$ represents the mean flow velocity normalized by V_s. Following Majdalani and Flandro,[34] we impose a constraint on the wave amplitude, namely,

$$\mathrm{M}^2 < \varepsilon_w < \mathrm{M}, \tag{11}$$

where $\mathrm{M} < 0.01$.

C. Leading-order decomposition

Equations (8)–(10) can be inserted back into Eqs. (6) and (7). The zeroth-order terms yield the mean flow equations

$$\nabla \cdot \mathbf{U} = 0, \tag{12}$$

$$\mathrm{M}^2(\mathbf{U} \cdot \nabla)\mathbf{U} = -\nabla p^{(0)} + \mathrm{M}^2\varepsilon[4\nabla(\nabla \cdot \mathbf{U})/3 - \nabla \times (\nabla \times \mathbf{U})]. \tag{13}$$

Following Berman,[5] a steady streamfunction can be defined by $\Psi = -xF(r)$. Subsequently, the velocity can be expressed by $(U_x, U_r) = [-xF'(r)/r, F(r)/r]$. By substituting these definitions into Eq. (13), Terrill has shown that $F = r^2$ provides the exact mean flow solution for the infinitely large suction case.[28] The mean pressure arising in this context can be integrated from Eq. (13) to obtain

$$p^{(0)}(x,r) = 1/\gamma - \mathrm{M}^2 r^2 (1 + x^2)/2. \tag{14}$$

D. Time-dependent equations

Terms of $\mathcal{O}(\varepsilon_w)$ in Eqs. (6) and (7) lead to

$$-i\omega\rho^{(1)} + \nabla \cdot \mathbf{u}^{(1)} = -\mathrm{M}\nabla \cdot (\rho^{(1)}\mathbf{U}), \tag{15}$$

$$-i\omega\mathbf{u}^{(1)} = -\mathrm{M}[\nabla(\mathbf{U} \cdot \mathbf{u}^{(1)}) - \mathbf{u}^{(1)} \times (\nabla \times \mathbf{U}) - \mathbf{U} \times (\nabla \times \mathbf{u}^{(1)})] - \nabla p^{(1)} + \mathrm{M}\varepsilon[4\nabla(\nabla \cdot \mathbf{u}^{(1)})/3 - \nabla \times (\nabla \times \mathbf{u}^{(1)})]. \tag{16}$$

Equations (15) and (16) describe the intimate coupling between mean and unsteady motions. They indicate that the wall suction velocity $\mathbf{U}$ can strongly influence the oscillatory flow motion.

IV. MOMENTUM TRANSPORT FORMULATION

A. Irrotational and solenoidal vectors

In order to proceed, temporal disturbances can be split into solenoidal and irrotational components.[35] Using a circumflex to denote the curl-free pressure-driven part, and a tilde for the divergence-free boundary-driven part, the time-dependent velocity component can be expressed as

$$\mathbf{u}^{(1)} = \hat{\mathbf{u}} + \tilde{\mathbf{u}}, \tag{17}$$

with

$$\nabla \times \mathbf{u}^{(1)} = \nabla \times \tilde{\mathbf{u}}, \quad p^{(1)} = \hat{p}, \quad \rho^{(1)} = \hat{\rho}. \tag{18}$$

This decomposition charges all vortices to the vortical field $\tilde{\mathbf{u}} = (\tilde{u}, \tilde{v})$, and compressibility sources and sinks to the acoustical field $\hat{\mathbf{u}} = (\hat{u}, \hat{v})$.

B. The linearized Navier-Stokes equations

Insertion of Eqs. (17) and (18) into Eqs. (15) and (16) leads to two independent sets that are coupled through the boundary conditions at the wall. These responses are byproducts of pressure-driven and vorticity-driven oscillation modes at $\mathcal{O}(\varepsilon_w)$. While the acoustical, compressible, and irrotational equations collapse into

$$-i\omega\hat{\rho} + \nabla \cdot \hat{\mathbf{u}} = -\mathrm{M}\nabla \cdot (\hat{\rho}\mathbf{U}), \tag{19}$$

$$-i\omega\hat{\mathbf{u}} = -\nabla\hat{p} + 4\mathrm{M}\varepsilon\nabla(\nabla\cdot\hat{\mathbf{u}})/3 - \mathrm{M}[\nabla(\hat{\mathbf{u}}\cdot\mathbf{U}) - \hat{\mathbf{u}}\times(\nabla\times\mathbf{U})], \tag{20}$$

the rotational and incompressible set comprises of

$$\nabla\cdot\tilde{\mathbf{u}} = 0, \tag{21}$$

$$-i\omega\tilde{\mathbf{u}} = -\mathrm{M}\varepsilon\nabla\times(\nabla\times\tilde{\mathbf{u}}) - \mathrm{M}[\nabla(\tilde{\mathbf{u}}\cdot\mathbf{U}) - \tilde{\mathbf{u}}\times(\nabla\times\mathbf{U}) - \mathbf{U}\times(\nabla\times\tilde{\mathbf{u}})]. \tag{22}$$

C. Coupling conditions

Two boundary conditions must be satisfied by the unsteady velocity component $\mathbf{u}^{(1)}$. These are the no-slip condition at the wall $u^{(1)}(x,1)=0$, and centerline symmetry, $\partial u^{(1)}(x,0)/\partial r = 0$.

D. Pressure-driven solution

When Eq. (19) is multiplied by $-i\omega$, the divergence of Eq. (20) can be evaluated; resulting terms can be added to produce the following wave equation:

$$\nabla^2\hat{p} + \omega^2\hat{p} = 4\mathrm{M}\varepsilon\nabla^2(\nabla\cdot\hat{\mathbf{u}})/3 - \mathrm{M}\{i\omega\nabla\cdot(\mathbf{U}\hat{p}) + \nabla^2(\hat{\mathbf{u}}\cdot\mathbf{U}) - \nabla\cdot[\hat{\mathbf{u}}\times(\nabla\times\mathbf{U})]\}. \tag{23}$$

At this juncture, a solution at $\mathcal{O}(\mathrm{M})$ can be achieved through separation of variables and closed-open boundary conditions. The ensuing acoustical pressure and velocity are

$$\hat{p} = \cos(\omega x) + \mathcal{O}(\mathrm{M}), \quad \hat{\mathbf{u}} = i\sin(\omega x)\mathbf{e}_x + \mathcal{O}(\mathrm{M}). \tag{24}$$

E. Vortical equations

Assuming that the ratio of the normal to axial velocity is of the same order as the Mach number [i.e. $\tilde{v}/\tilde{u} = \mathcal{O}(\mathrm{M})$], $\tilde{v}$ can be dropped at leading order. This assumption can be justified in view of the arguments presented by Flandro.[36] Applying this condition, along with the definition of the mean flow velocity, the axial momentum equation reduces to

$$i\mathrm{Sr}\tilde{u} = \frac{\partial(\tilde{u}\mathbf{U})}{\partial x} + v_0\frac{\partial\tilde{u}}{\partial r} - \frac{\varepsilon}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\tilde{u}}{\partial r}\right) + \mathcal{O}(\mathrm{M}), \tag{25}$$

where $\mathrm{Sr} = \omega/\mathrm{M}$ is the Strouhal number. When expressed in terms of the mean flow streamfunction, Eq. (25) becomes

$$\left(i\mathrm{Sr} + \frac{F'}{r}\right)\tilde{u} = \frac{F}{r}\frac{\partial\tilde{u}}{\partial r} - \frac{xF'}{r}\frac{\partial\tilde{u}}{\partial x} - \frac{\varepsilon}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\tilde{u}}{\partial r}\right) + \mathcal{O}(\mathrm{M}). \tag{26}$$

A solution for Eq. (26) will be presented next.

F. The separable boundary-layer equation

A solution for Eq. (26) can be developed through the use of separation of variables. Assuming the form

$$\tilde{u}(x,r) = X(x)Y(r), \tag{27}$$

substitution into Eq. (26) leads to

$$\frac{x}{X}\frac{\mathrm{d}X}{\mathrm{d}x} = \frac{F}{F'Y}\frac{\mathrm{d}Y}{\mathrm{d}r} - \frac{\varepsilon r}{F'Y}\frac{\mathrm{d}^2Y}{\mathrm{d}r^2} - \frac{\varepsilon}{F'Y}\frac{\mathrm{d}Y}{\mathrm{d}r} - \frac{ir\mathrm{Sr}}{F'} - 1 = \kappa_n, \tag{28}$$

where $\kappa_n > 0$ is the separation eigenvalue. Integration of the x equation can be performed easily and then inserted into Eq. (27). The outcome is

$$\tilde{u}(x,r) = \sum_n c_n x^{\kappa_n} Y_n(r), \tag{29}$$

where c_n is the integration constant for each κ_n. Satisfaction of the no-slip condition at the wall requires setting the acoustical and vortical velocity components equal and opposite at $r=1$. One finds

$$\tilde{u}(x,1) = -i\sin(\omega x). \tag{30}$$

Using a series expansion of the sine function, and setting the result equal to Eq. (29), one gets

$$\sum_n c_n x^{\kappa_n} Y_n(1) = -i\sum_{n=0}^{\infty}\frac{(-1)^n(\omega x)^{2n+1}}{(2n+1)!}. \tag{31}$$

Equating terms necessitates that

$$\kappa_n = 2n+1, \quad c_n = -i\frac{(-1)^n\omega^{2n+1}}{(2n+1)!}, \quad Y_n(1) = 1, \tag{32}$$

where $n = 0,1,2,\ldots,\infty$. The rotational velocity component becomes

$$\tilde{u}(x,r) = -i\sum_{n=0}^{\infty}\frac{(-1)^n(\omega x)^{2n+1}}{(2n+1)!}Y_n(r). \tag{33}$$

In order to bring closure to Eq. (33), Y_n needs to be determined from Eq. (28). One finds that Y_n must be obtained from the doubly perturbed boundary-value problem, namely,

$$\varepsilon\frac{\mathrm{d}^2Y_n}{\mathrm{d}r^2} + [(\varepsilon - F)/r]\frac{\mathrm{d}Y_n}{\mathrm{d}r} + [i\mathrm{Sr} + (2n+2)F'/r]Y_n = 0, \tag{34}$$

where

$$Y_n(1) = 1, \quad Y_n'(0) = 0. \tag{35}$$

These two boundary conditions stem from the no-slip and core symmetry requirements.

V. BOUNDARY-LAYER ANALYSIS

Substitution of Terrill's mean flow solution $F = r^2$ into Eq. (34) leads to

$$\varepsilon\frac{\mathrm{d}^2Y_n}{\mathrm{d}r^2} + (-r + \varepsilon/r)\frac{\mathrm{d}Y_n}{\mathrm{d}r} + (i\mathrm{Sr} + 4n + 4)Y_n = 0. \tag{36}$$

In what follows, Eq. (36) will be solved using the method of matched-asymptotic expansions. To that end, the perturbation parameters need to be first identified. Since our concern is with solutions corresponding to large Re, the primary perturbation parameter is clearly $\varepsilon = \mathrm{Re}^{-1} \ll 1$. Furthermore, one must recognize that the condition of $\mathrm{Sr} \gg 1$ is necessary to ensure a sufficiently oscillatory flow. It may be instructive to note that, according to (25), the Strouhal number depends on the product of the circular frequency and tube radius ($\omega_0 a$) divided by the suction velocity V_s. Since V_s is usually two to three orders of magnitude smaller than the speed of sound, Sr is about two to three orders of magnitude

larger than the aeroacoustic Strouhal number based on a_0 instead of V_s. Recalling that the latter extends over $[10^{-3},10]$ with a peak in the noise spectrum at $\mathrm{Sr}\approx 0.2$, the current Strouhal number extends over the range $[1,10^3]$ with typically reported values of $\mathrm{Sr}\approx 20-50$.

Asymptotic approximations to Eq. (36) depend on the development of a relationship between the two perturbation parameters present in the problem. By inspection of numerical simulations carried out for the large suction case, one comes to the conclusion that the problem exhibits a typical second-order wave-type response that bears a strong resemblance to a critically damped wave. On that account, a distinguished limit will be needed to relate ε and Sr in a manner to produce the expected response.

To start, an order of magnitude relationship between the control parameters must be posited. Without loss in generality, one can let

$$\mathrm{Sr}\sim\mathcal{O}(\varepsilon^{-\zeta}). \tag{37}$$

Subsequently, rescaling of the viscous domain requires a distortion of the independent variable in the form

$$1-r=\varepsilon^k z. \tag{38}$$

In order to determine the distinguished limit, one may apply the stretching transformation and use $\mathrm{Sr}=\varepsilon^{-\zeta}$ in Eq. (36). The result is

$$\varepsilon^{1-2k}\frac{\mathrm{d}^2Y_n}{\mathrm{d}z^2}+\varepsilon^{-k}(r-\varepsilon/r)\frac{\mathrm{d}Y_n}{\mathrm{d}z}+[i\varepsilon^{-\zeta}+(4n+4)]Y_n=0. \tag{39}$$

For a critically damped response to occur near the wall, a balance between all three terms in Eq. (39) must be established. Clearly, all terms will be in balance when $\zeta=k=1$. These distinctive orders indicate that the boundary-layer thickness is of $\mathcal{O}(\varepsilon)$ and that

$$\mathrm{Sr}=\mathcal{O}(\varepsilon^{-1}). \tag{40}$$

It may be interesting to note that these distinguished limits are dissimilar from those realized in the injection flow analog,[37–42] including those arising in the rectangular cavity.[1,2] The disparity can be attributed to the reversal in the physics of the problem, namely, in the relocation of the viscous boundary layer to the vicinity of the suction wall.

VI. MATCHED-ASYMPTOTIC EXPANSIONS

A. The relevant scales

In order to proceed, one has to identify the length scale needed to magnify the thin viscous region near the wall. From the foregoing order of magnitude analysis, the relevant scales can be recognized to be $r=r$ in the outer domain and

$$z=(1-r)/\varepsilon \tag{41}$$

in the inner region. Solving the problem with matched-asymptotic expansions involves the formulation of two separate solutions over the domain of interest. While Eq. (36) is only valid in the outer domain (i.e., the inviscid region), a transformed equation is needed to capture the rapid variations near the wall (inside the viscous boundary layer). In both cases, we find it convenient to multiply Eq. (36) by r and write the governing equation as

$$\varepsilon r\frac{\mathrm{d}^2Y_n}{\mathrm{d}r^2}+(-r^2+\varepsilon)\frac{\mathrm{d}Y_n}{\mathrm{d}r}+r[i\mathrm{Sr}+(4n+4)]Y_n=0. \tag{42}$$

B. The outer solution

Using regular perturbations to construct the outer solution Y_n^o, one may start by putting

$$Y_n^o=Y_0^o+\varepsilon Y_1^o+\mathcal{O}(\varepsilon^2). \tag{43}$$

Inserting Eq. (43) into Eq. (42) gives

$$\varepsilon r\frac{\mathrm{d}^2Y_0^o}{\mathrm{d}r^2}-r^2\frac{\mathrm{d}Y_0^o}{\mathrm{d}r}-r^2\varepsilon\frac{\mathrm{d}Y_1^o}{\mathrm{d}r}+\varepsilon\frac{\mathrm{d}Y_0^o}{\mathrm{d}r}+r[(4n+4)+i\mathrm{Sr}]\times(Y_0^o+\varepsilon Y_1^o)+\mathcal{O}(\varepsilon^2)=0. \tag{44}$$

Keeping in mind that $\mathrm{Sr}=\mathcal{O}(\varepsilon^{-1})$, the equations defining the first two terms in the outer solution become

$$i\mathrm{Sr}Y_0^o=0, \tag{45}$$

$$i\mathrm{Sr}\varepsilon Y_1^o=r^2\frac{\mathrm{d}Y_0^o}{\mathrm{d}r}-r(4n+4)Y_0^o. \tag{46}$$

Solving these equations leads to

$$Y_0^o=Y_1^o=0,\quad Y_n^o=0+\mathcal{O}(\varepsilon^2). \tag{47}$$

C. The inner solution

Having realized that the outer solution is zero, the stretching transformation must now be applied to the original coordinate in order to obtain the inner equation. This procedure converts Eq. (42) into

$$(1-\varepsilon z)\frac{\mathrm{d}^2Y_n^i}{\mathrm{d}z^2}+[1-\varepsilon(2z+1)+\varepsilon^2z]\frac{\mathrm{d}Y_n^i}{\mathrm{d}z}+(\varepsilon-\varepsilon^2z)\times[(4n+4)+i\mathrm{Sr}]Y_n^i=0. \tag{48}$$

The inner solution can be similarly expanded using $Y_n^i=Y_0^i+\varepsilon Y_1^i+\mathcal{O}(\varepsilon^2)$. The outcome is

$$(1-\varepsilon z)\frac{\mathrm{d}^2Y_0^i}{\mathrm{d}z^2}+(\varepsilon-\varepsilon^2z)\frac{\mathrm{d}^2Y_1^i}{\mathrm{d}z^2}+[1-\varepsilon(2z+1)+\varepsilon^2z]\times\left(\frac{\mathrm{d}Y_0^i}{\mathrm{d}z}+\varepsilon\frac{\mathrm{d}Y_1^i}{\mathrm{d}z}\right)+(\varepsilon-\varepsilon^2z)[(4n+4)+i\mathrm{Sr}]\times(Y_0^i+\varepsilon Y_1^i)=0. \tag{49}$$

Since the inner equation is of second order, two conditions must be imposed on the inner solution at each perturbation level. While the first can be determined from the no-slip at the wall, the second must be deduced by matching with the outer domain. Using Eq. (35) and the expansion for Y_n^i, the boundary condition at the wall gives

$$Y_0^i(z=0)=1,\quad Y_1^i(z=0)=0. \tag{50}$$

At this juncture, the leading and first-order correction terms can be readily found. From Eq. (49), the $\mathcal{O}(1)$ equation reads

$$\frac{d^2Y_0^i}{dz^2}+\frac{dY_0^i}{dz}+i\mathrm{Sr}\varepsilon Y_0^i=0, \tag{51}$$

wherefore

$$Y_0^i=c_1\exp[\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1)z] + c_2\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)z]. \tag{52}$$

Straightforward application of the boundary condition at the wall yields

$$c_2=1-c_1, \tag{53}$$

so that

$$Y_0^i=c_1\exp[\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1)z] + (1-c_1)\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)z]. \tag{54}$$

Next, the $\mathcal{O}(\varepsilon)$ equation can be collected from Eq. (49). One obtains

$$\frac{d^2Y_1^i}{dz^2}+\frac{dY_1^i}{dz}+i\mathrm{Sr}\varepsilon Y_1^i=z\frac{d^2Y_0^i}{dz^2}+(2z+1)\frac{dY_0^i}{dz} -(4n+4-i\mathrm{Sr}\varepsilon z)Y_0^i. \tag{55}$$

While the homogeneous solution can be evaluated by inspection via

$$Y_{1,\mathrm{h}}^i=B_1\exp[\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1)z] + B_2\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)z], \tag{56}$$

the right-hand side of Eq. (55) can be rearranged into

$$c_1[\tfrac{1}{4}(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1)^2z+(z+\tfrac{1}{2})(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1) -(4n+4-i\mathrm{Sr}\varepsilon z)]\exp[\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1)z] +(1-c_1)[\tfrac{1}{4}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)^2z-(z+\tfrac{1}{2})(\sqrt{1-4i\mathrm{Sr}\varepsilon} +1)-(4n+4-i\mathrm{Sr}\varepsilon z)\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)z]. \tag{57}$$

A particular solution must therefore be assumed such that

$$Y_{1,\mathrm{p}}^i=(B_3z+B_4z^2)\exp[\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1)z] + (B_5z+B_6z^2)\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)z]. \tag{58}$$

After differentiating and substituting Eq. (58) into the left-hand side of Eq. (55), equating terms of order 1 and z^2 requires that

$$B_3=c_1\left[\frac{1}{2}-\frac{(4n+\frac{9}{2})}{\sqrt{1-4i\mathrm{Sr}\varepsilon}}\right],\quad B_4=0,$$

$$B_5=(1-c_1)\left[\frac{1}{2}+\frac{(4n+\frac{9}{2})}{\sqrt{1-4i\mathrm{Sr}\varepsilon}}\right],\quad B_6=0. \tag{59}$$

By writing $Y_1^i=Y_{1,\mathrm{h}}^i+Y_{1,\mathrm{p}}^i$ and enforcing Eq. (50), the inner solution turns into

$$Y_n^i=\left\{c_1-B_2\varepsilon+c_1z\varepsilon\left[\frac{1}{2}-\frac{(4n+\frac{9}{2})}{\sqrt{1-4i\mathrm{Sr}\varepsilon}}\right]\right\} \times\exp[\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}-1)z] +\left\{(1-c_1)+B_2\varepsilon+(1-c_1)z\varepsilon\left[\frac{1}{2}+\frac{(4n+\frac{9}{2})}{\sqrt{1-4i\mathrm{Sr}\varepsilon}}\right]\right\} \times\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)z]. \tag{60}$$

D. Asymptotic matching

Inner and outer solutions can be readily matched using Prandtl's matching principle.[43] By requiring the inner solution in the outer domain to match the outer solution in the inner domain, one may set

$$Y_n^i(z\to\infty)=Y_n^o(r\to0)=Y_{n,\mathrm{cp}}, \tag{61}$$

where $Y_{n,\mathrm{cp}}$ represents the common part in the overlap region shared by both inner and outer solutions. In our problem, both the outer and common parts are zero. The inner solution in the outer domain will also vanish according to Eq. (61) if, and only if, $c_1=B_2=0$. These constants bring closure to the inner solution by permitting the construction of a uniformly valid composite solution. Hence, by adding the inner and outer solutions, less $Y_{n,\mathrm{cp}}$, one finally obtains

$$Y_n(r)=\{1+(1-r)[\tfrac{1}{2}+(4n+\tfrac{9}{2})/\sqrt{1-4i\mathrm{Sr}\varepsilon}]\} \times\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)(1-r)/\varepsilon]. \tag{62}$$

E. The oscillatory velocity

Insertion of Eq. (62) into Eq. (33) results in an expression for the rotational velocity component. The addition of the acoustical component, given by Eq. (24), enables us to express the total axial velocity as an infinite sum, namely,

$$u^{(1)}(x,r)=i\left(\sin(\omega x)-\sum_{n=0}^{\infty}\frac{(-1)^n(\omega x)^{2n+1}}{(2n+1)!} \times\left\{1+(1-r)\left[\frac{1}{2}+\frac{(4n+\frac{9}{2})}{\sqrt{1-4i\mathrm{Sr}\varepsilon}}\right]\right\} \times\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)(1-r)/\varepsilon]\right). \tag{63}$$

Since $(1-r)$ is small near the wall, one may use $n=0$ in the secondary term arising from the first-order inner correction. The resulting expression can be summed, at leading order, over all eigenvalues, and placed in closed form by recognizing and grouping the implicit sine function expansion. This manipulation produces

$$u^{(1)}(x,r)=i\sin(\omega x)\{1-[1+\tfrac{1}{2}(1-r)(1+9/\sqrt{1-4i\mathrm{Sr}\varepsilon})] \times\exp[-\tfrac{1}{2}(\sqrt{1-4i\mathrm{Sr}\varepsilon}+1)(1-r)/\varepsilon]\}. \tag{64}$$

Being practically equivalent to Eq. (63), this formula completes our derivation of the acoustico-vortical contribution. Equation (64) clearly displays the key parameters affecting the unsteady wave behavior.

The foregoing methodology sketches the procedural steps needed to arrive at a field approximation based on a matched-asymptotic solution of the vorticity wave. The same approach can be employed in other suction-driven problems exhibiting more sophisticated mean flow functions. For appreciable suction, however, an exact solution of the rotational contribution is possible. This will be presented next.

VII. EXACT SOLUTION

A. Motivation

For the case of $F=r^2$, Eq. (34) can be solved exactly. This requires setting

$$X=\phi, \quad B(X)=\psi Y_n, \tag{65}$$

where ϕ and ψ are functions of the radial coordinate r. These transformations yield

$$\frac{\mathrm{d}Y_n}{\mathrm{d}r}=-\frac{\psi'}{\psi^2}B+\frac{\phi'}{\psi}\frac{\mathrm{d}B}{\mathrm{d}X}, \quad \frac{\mathrm{d}^2Y_n}{\mathrm{d}r^2}=\frac{\phi'^2}{\psi}\frac{\mathrm{d}^2B}{\mathrm{d}X^2}+\left(\frac{\phi''}{\psi}-\frac{2\phi'\psi'}{\psi^2}\right)\frac{\mathrm{d}B}{\mathrm{d}X}-\left(\frac{\psi''}{\psi^2}-\frac{2\psi'^2}{\psi^3}\right)B, \tag{66}$$

where primes stand for differentiation with respect to r. Substitution into Eq. (34) gives

$$\frac{\mathrm{d}^2B}{\mathrm{d}X^2}+\frac{1}{\phi'^2}\left(\phi''-\frac{2\phi'\psi'}{\psi}-r\,\mathrm{Re}\,\phi'+\frac{\phi'}{r}\right)\frac{\mathrm{d}B}{\mathrm{d}X}+\left[-\frac{\psi''}{\psi}+\frac{2\psi'^2}{\psi^2}+\frac{r\,\mathrm{Re}\,\psi'}{\psi}-\frac{\psi'}{r\psi}+\mathrm{Re}(i\mathrm{Sr}+4n+4)\right]\frac{B}{\phi'^2}=0. \tag{67}$$

At this point, ψ and ϕ are chosen so that the variable coefficients in Eq. (67) are turned into pure constants. For that purpose, the coefficient of the first derivative is suppressed via

$$\phi''-2\phi'\psi'/\psi-r\,\mathrm{Re}\,\phi'+\phi'/r=0,$$

$$\psi'/\psi=\tfrac{1}{2}(\phi''/\phi'-r\,\mathrm{Re}+r^{-1}). \tag{68}$$

Integrating Eq. (68) gives $\psi=T_0\sqrt{r\phi'}\exp(-\mathrm{Re}\,r^2/4)$, where T_0 is constant. Equation (67) simplifies into

$$B_{XX}+[\mathrm{Re}(i\mathrm{Sr}+4n+4)\phi'^{-2}+\beta]B=0, \tag{69}$$

where

$$\beta=\frac{1}{\phi'^2}\left[-\frac{\psi''}{\psi}+\frac{2\psi'^2}{\psi^2}+\left(r\,\mathrm{Re}-\frac{1}{r}\right)\frac{\psi'}{\psi}\right]. \tag{70}$$

By setting $\mathrm{Re}(i\mathrm{Sr}+4n+4)/\phi'^2=\mathrm{const}$, one obtains $\phi'=\sqrt{\mathrm{Re}}$ and $X=\phi=r\sqrt{\mathrm{Re}}$. Without losing generality, one puts $T_0=1/\sqrt[4]{\mathrm{Re}}$ so that $\psi(r)=\sqrt{r}\exp(-\mathrm{Re}\,r^2/4)$. The outcome is

$$B_{XX}+[\sigma-\tfrac{1}{4}(X^2-1/X^2)]B=0; \quad \sigma=5+4n+iS \tag{71}$$

B. Complete solution

At this point, two additional variable transformations are needed. The first applies to the dependent variable via $H(X)=B(X)\sqrt{X}$. Equation (71) becomes

$$H_{XX}-H_X/X+(\sigma-\tfrac{1}{4}X^2+1/X^2)H=0. \tag{72}$$

The second transformation, namely, $Z=X^2/2$, affects the independent variable. This results in

$$\frac{\mathrm{d}^2H}{\mathrm{d}Z^2}+\left(-\frac{1}{4}+\frac{\sigma}{2Z}+\frac{1}{4Z^2}\right)H=0, \tag{73}$$

whose solution can be written as

$$H(Z)=C_1M_{(1/2)\,\sigma,0}(Z)+C_2W_{(1/2)\,\sigma,0}(Z), \tag{74}$$

where C_1 and C_2 are constants to be determined from boundary conditions; the functions M and W, on the other hand, are the Whittaker functions of the first and second kind. The Whittaker functions are related to the Kummer or confluent hypergeometric functions through[44]

$$M_{\kappa,\mu}(Z)=Z^{(1/2)+\mu}e^{-(1/2)Z}\Phi(\tfrac{1}{2}+\mu-\kappa,1+2\mu;Z),$$

$$W_{\kappa,\mu}(Z)=Z^{(1/2)+\mu}e^{-(1/2)Z}\Psi(\tfrac{1}{2}+\mu-\kappa,1+2\mu;Z), \tag{75}$$

wherefrom

$$H(Z)=\sqrt{Z}\exp(-\tfrac{1}{2}Z)[C_1\Phi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;Z)+C_2\Psi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;Z)]. \tag{76}$$

Returning to original variables, one gets

$$Y_n(r)=\sqrt[4]{\mathrm{Re}/4}[C_1\Phi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;\tfrac{1}{2}\mathrm{Re}\,r^2)+C_2\Psi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;\tfrac{1}{2}\mathrm{Re}\,r^2)]. \tag{77}$$

Note that $\sqrt[4]{\mathrm{Re}/4}$ is a constant that can be absorbed into C_1 and C_2. Thus, without loss in generality, one puts

$$Y_n(r)=C_1\Phi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;\tfrac{1}{2}\mathrm{Re}\,r^2)+C_2\Psi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;\tfrac{1}{2}\mathrm{Re}\,r^2),$$

$$Y_n'(r)=r\,\mathrm{Re}(\tfrac{1}{2}-\tfrac{1}{2}\sigma)[C_1\Phi(\tfrac{3}{2}-\tfrac{1}{2}\sigma,2;\tfrac{1}{2}\mathrm{Re}\,r^2)-C_2\Psi(\tfrac{3}{2}-\tfrac{1}{2}\sigma,2;\tfrac{1}{2}\mathrm{Re}\,r^2)]. \tag{78}$$

To find C_2, $\Psi(a,b;z)$ must be expressed in terms of $\Phi(a,b;z)$ using[44]

$$\Psi(a,b;x)=\frac{\pi}{\sin(\pi b)}\left[\frac{\Phi(a,b;x)}{\Gamma(1+a-b)\Gamma(b)}-x^{1-b}\frac{\Phi(1+a-b,2-b;x)}{\Gamma(a)\Gamma(2-b)}\right]. \quad (79)$$

This expression leads to an infinite value at the core for $b>1$ except when $C_2=0$. The remaining C_1 is determined using the boundary condition at the wall. One finds

$$Y_n(r)=\Phi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;\tfrac{1}{2}\mathrm{Re}\,r^2)/\Phi(\tfrac{1}{2}-\tfrac{1}{2}\sigma,1;\tfrac{1}{2}\mathrm{Re}), \quad (80)$$

which indicates that the characteristic coordinate scales with $r\sqrt{\mathrm{Re}/2}$. Following the same lines as before, one can sum the acoustical and vortical contributions to obtain the unsteady velocity component. This takes the form

$$u^{(1)}(x,r)=i\left[\sin(\omega x)-\sum_{n=0}^{\infty}\frac{(-1)^n(\omega x)^{2n+1}\Phi(-2n-2-\frac{1}{2}i\mathrm{Sr},1;\frac{1}{2}\mathrm{Re}\,r^2)}{(2n+1)!\Phi(-2n-2-\frac{1}{2}i\mathrm{Sr},1;\frac{1}{2}\mathrm{Re})}\right]. \quad (81)$$

Unlike the matched-asymptotic solution given by Eq. (62), the physical clarity of the exact solution is encumbered by the infinite summation of the Kummer function.

VIII. DISCUSSION

A. A formerly tested methodology

The decomposition of the time-dependent governing equations presented in Sec. IV, during the momentum transport formulation, was first introduced by Flandro[42] and further developed by Majdalani and Roh.[38] For injection-driven flows with superimposed oscillations, a similar analytical framework has provided accurate flow field predictions. Those could be substantiated using fully nonlinear computational models. The asymptotic approximations obtained previously were also shown to agree favorably with experimental data obtained in cold-flow simulations of transpiring surfaces.[39–42] Although the physical nature of the problem changes when suction is introduced, the assumptions used in reducing the governing equations remain valid, regardless of the inflow or outflow boundary conditions. By analogy with the injection-driven problem, one can expect a comparable level of agreement between the asymptotic formulations given here and either numerical or experimental studies of the model at hand. However, in the absence of experimental measurements to compare with, numerical simulations must be resorted to.

B. Numerical confirmation

Due to the lack of experimental data for the suction-driven flow, we have compared the matched-asymptotic expansion to a numerical solution of Eq. (36) obtained from a code that was originally developed for injection-driven flows by Majdalani and Van Moorhem.[39] The same code was tested by comparing its results to the exact solution derived in Sec. VII. The algorithm employs a fixed step fifth-order Runge-Kutta method with shooting to handle the boundary conditions. For the suction case, the step size used was 10^{-6} on a normalized interval. In former studies,[39–41] the same code was shown to provide satisfactory agreement with experimental data. Therein, the code was also shown to closely match computational data obtained independently by Yang and Roh who utilized a fully compressible, finite-volume Navier-Stokes solver.[45]

C. Graphical confirmation

Figure 2 illustrates the agreement between the exact or numerical solutions and Eq. (62). Over typical ranges of physical parameters, the graphical comparison clearly indicates that the matched-asymptotic solution is in close agreement with the numerical results. Graphically, the accuracy of the approximate formulation is seen to increase with increasing Reynolds and Strouhal numbers. This observation is reassuring since it indicates that the solution exhibits the proper asymptotic behavior as $\varepsilon=\mathrm{Re}^{-1}\to 0$ and $\mathrm{Sr}^{-1}\to 0$. It is also satisfying to note the nearly critical damped-wave response. This rapid damping in both depth and amplitude is consistent with the arguments introduced in Sec. V regarding the scaling orders of Sr and ε. Also note that the wave behavior is different from the highly under-damped wave solution associated with injection-driven flows. In the latter, numerous peaks of diminishing amplitude appear as the distance from the wall is increased.[1]

In order to assess the truncation error associated with Eq. (62), the maximum absolute error between asymptotics and numerics is shown in Fig. 3 for the first three eigenvalues and a range of Sr and Re. When plotted versus ε, this error is seen to exhibit a clear asymptotic order as $\varepsilon\to 0$. It also decreases in magnitude with successive increases in Sr. It can also be seen that the slope of the error curves and, in turn, the order of the truncation error approach unity for

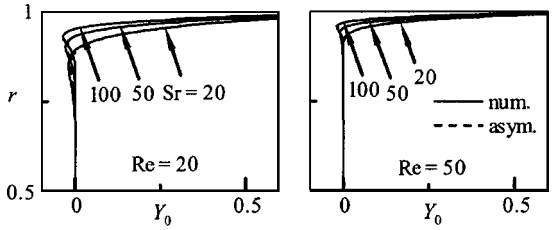


FIG. 2. Here Y_n is plotted for $n=0$ over a range of Reynolds and Strouhal numbers. The figures show the slightly under-damped response at Re=20 and 50.

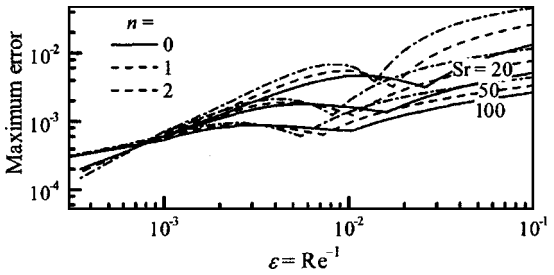


FIG. 3. Plot of the maximum error in the approximate solution versus ε.

sufficiently small ε. This confirms the order claimed for this approximation. In particular, it may be useful to note that the slight increase in the error intercept at higher eigenvalues does not affect the summed up solution. This is due to the rapid convergence of the series in Eq. (63) as n is increased.

D. Variation of flow parameters

The four subsets of Fig. 4 illustrate the effect of varying either the suction velocity or the oscillation frequency on the time-dependent solution. In all four cases, the velocity is seen to be a wave traveling in time. While a viscous and rotational layer is present near the wall, a broad inviscid and irrotational region covers the remaining domain. Interestingly, the unsteady velocity reaches a maximum value inside the viscous layer where a small velocity overshoot is realized near the wall. This phenomenon is well known as Richardson's annular effect and seems to be characteristic of oscillatory flows in tubes and channels with and without wall permeation.[46] The small percentage overshoot that accompanies a suction-driven flow is of the same order as that associated with the exact Sexl profile inside a nonporous tube. It is significantly smaller than the 100% overshoot (i.e., velocity doubling) that recurs near the walls of injection-driven flows.

According to the theory of laminar periodic flows, one could expect the magnitude of the velocity overshoot to increase at higher oscillation frequencies.[46] The reason is this. As the Strouhal number is increased, the spatial wavelength diminishes, being inversely proportional to Sr. The first oscillation peak stemming from a favorable coupling between acoustical and vortical waves will then form closer to the wall. Since the rotational component diminishes with the distance from the wall, a larger vortical contribution augments the acoustical wave when their coupling occurs closer to the wall (e.g., at higher frequencies). The reduction in spatial wavelength at higher Strouhal numbers increases the rate of viscous dissipation and causes the boundary-layer thickness to decrease.[47] The latter is often referred to as the penetration depth and is a measure of the viscous and rotational layer above the solid boundary.

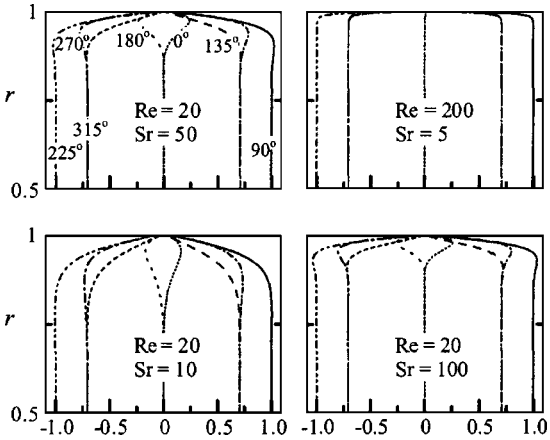


FIG. 4. The oscillatory velocity $u^{(1)}\exp(-it)$ is shown at four different times for $m=1$ and $x/l=0.5$. Angles in the figures represent dimensionless time. Properties correspond to an order of magnitude variation in Reynolds and Strouhal numbers.

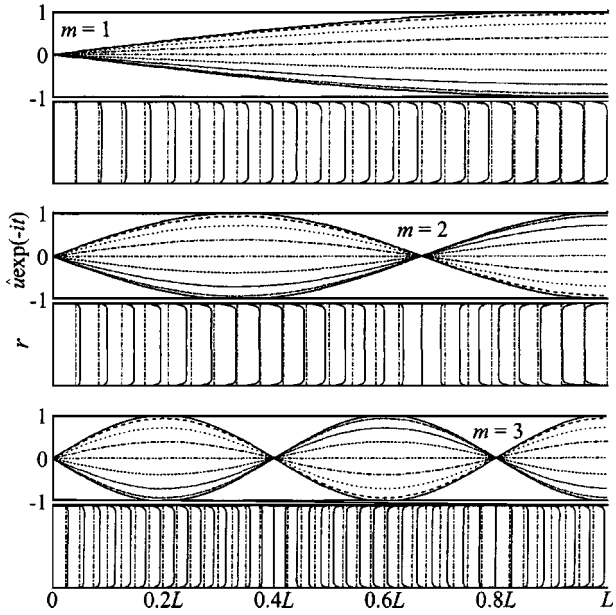


FIG. 5. The spatial distribution of the oscillatory velocity $u^{(1)}\exp(-it)$ is illustrated for the first three acoustical oscillation modes by plotting the velocity modulus at several evenly spaced axial positions for Re=20 and Sr=10. The corresponding longitudinal mode shapes of the acoustical velocity component are also shown at nine equally dispersed times. In addition to the head end location, the smallest disturbances occur at the nth internal acoustical node located at $x^*/L=n/(m-\frac{1}{2})$, $n<m$, $m=1,2,3$, etc.

The top two subsets in Fig. 4 illustrate the effects of increasing the Reynolds number while decreasing the Strouhal number via an order of magnitude increase in V_s. As the suction speed is increased, the rotational layer is reduced in both depth and overshoot. While the reduction in overshoot can be attributed to the smaller vortical contribution associated with a smaller Sr, the smaller depth may be attributed to the increased Re. Evidently, the increased fluid withdrawal rate has the effect of attracting the viscous layer closer to the wall.

The bottom two subsets, on the other hand, confirm the previous statements made regarding the oscillation frequency. Clearly, through an order of magnitude increase in Sr, the penetration depth is decreased, while Richardson's effect is made more appreciable.

E. Oscillation modes

In closing, we use Fig. 5 to show the spatial evolution of the oscillatory velocity for the first three oscillation modes. Also plotted are the amplitudes of the inviscid velocity at nine equally scattered times. This is done to illustrate the

strong correspondence between the pressure-driven inviscid mode shapes and the spatial distribution of the total velocity. Since the rotational contribution always decays away from the walls, it is clear that the inviscid solution dominates near the core. The spatial amplitude of the oscillatory velocity is hence controlled by the pressure-driven mode shapes associated with the inviscid wave. Except for the small viscous layer that is drawn to the wall by hard suction, the flow is primarily irrotational. In addition to the head end location where no oscillations can be entertained, the weakest disturbances take place in the vicinity of nth internal acoustical nodes. This is due to the smallest wave amplitudes being located at $x/l=n/(m-\frac{1}{2})$ for all $n<m$ as shown on the graph.

In this article, the main thrust has been placed on presenting the procedural steps needed to obtain an unsteady flow approximation based on matched-asymptotic expansions of the vortical wave contribution. For the special case of appreciable wall suction, an exact solution was also possible. Despite its algebraic opacity, the exact solution served a dual purpose by helping to validate both numerics and asymptotics. In later work, we hope to extend this analysis to an oscillating gas with arbitrary levels of suction.

ACKNOWLEDGMENT

The authors are deeply grateful for the support received from the National Science Foundation under Grant No. CMS-0353518.

[1] J. Majdalani, "Vortical and acoustical mode coupling inside a two-dimensional cavity with transpiring walls," J. Acoust. Soc. Am. **106**, 46–56 (1999).
[2] J. Majdalani, "Improved solution for the vortical and acoustical mode coupling inside a two-dimensional cavity with porous walls," J. Acoust. Soc. Am. **109**, 475–479 (2001).
[3] R. G. Gutman, *Membrane Filtration* (Adam Hilger, Bristol, UK, 1987).
[4] V. Nassehi, "Modelling of combined Navier-Stokes and Darcy flows in crossflow membrane filtration," Chem. Eng. Sci. **53**(6), 1253–1265 (1998).
[5] A. S. Berman, "Laminar flow in channels with porous walls," J. Appl. Phys. **24**(9), 1232–1235 (1953).
[6] A. S. Berman, "Effects of porous boundaries on the flow of fluids in systems with various geometries," Proceedings of the Second United Nations International Conference on the Peaceful Uses of Atomic Energy (1958), Vol. 4, pp. 351–358.
[7] A. S. Berman, "Laminar flow in an annulus with porous walls," J. Appl. Phys. **29**(1), 71–75 (1958).
[8] J. M. Drazen, R. D. Kamm, and A. S. Slutsky, "High frequency ventilation," Physiol. Rev. **64**(2), 505–543 (1984).
[9] C. Y. Wang, "Pulsatile flow in a porous channel," Trans. AMSE: Ser. E J. Appl. Mech. **38**, 553–555 (1971).
[10] Y. C. Fung and C. S. Yih, "Peristaltic transport," J. Appl. Mech. **35**, 669–675 (1968).
[11] H. N. Chang, J. S. Ha, J. K. Park, I. H. Kim, and H. D. Shin, "Velocity field of pulsatile flow in a porous tube," J. Biomech. **22**(11-12), 1257–1262 (1989).
[12] G. D. Raithby and D. C. Knudsen, "Hydrodynamic development in a duct with suction and blowing," Trans. ASME: Ser. E **41**, 896–902 (1974).
[13] J. P. Hartnell and E. R. G. Eckert, "Mass-transfer cooling in a laminar boundary layer with constant fluid properties," Trans. ASME **79**, 247–254 (1957).
[14] G. Avalon, G. Casalis, and J. Griffond, "Flow instabilities and acoustic resonance of channels with wall injection," presented at the AIAA 98-3218, Cleveland, OH, 1998.
[15] G. Casalis, G. Avalon, and J.-P. Pineau, "Spatial instability of planar channel flow with fluid injection through porous walls," Phys. Fluids **10**(10), 2558–2568 (1998).
[16] J. Griffond, G. Casalis, and J.-P. Pineau, "Spatial instability of flow in a semiinfinite cylinder with fluid injection through its porous walls," Eur. J. Mech. B/Fluids **19**(1), 69–87 (2000).
[17] M. B. Zaturska, P. G. Drazin, and W. H. H. Banks, "On the flow of a viscous fluid driven along a channel by suction at porous walls," Fluid Dyn. Res. **4**(3), 151–178 (1988).
[18] C. L. Taylor, W. H. H. Banks, M. B. Zaturska, and P. G. Drazin, "Three-dimensional flow in a porous channel," Q. J. Mech. Appl. Math. **44**(1), 105–133 (1991).
[19] W. H. H. Banks and M. B. Zaturska, "On flow through a porous annular pipe," Phys. Fluids **4**(6), 1131–1141 (1992).
[20] M. B. Zaturska and W. H. H. Banks, "Suction-driven flow in a porous pipe," J. Appl. Math. Mech. **75**(1), 21–30 (1995).
[21] W. A. Robinson, "The existence of multiple solutions for the laminar flow in a uniformly porous channel with suction at both walls," J. Eng. Math. **10**(1), 23–40 (1976).
[22] K.-G. Shih, "On the existence of solutions of an equation arising in the theory of laminar flow in a uniformly porous channel with injection," SIAM J. Appl. Math. **47**(3), 526–533 (1987).
[23] S. P. Hastings, C. Lu, and A. D. MacGillivray, "A boundary value problem with multiple solutions from the theory of laminar flow," SIAM J. Math. Anal. **23**(1), 201–208 (1992).
[24] C. Lu, "On the existence of multiple solutions of a boundary value problem arising from laminar flow through a porous pipe," Cana. Appl. Mathematics Q. **2**(3), 361–393 (1994).
[25] S. W. Yuan and A. B. Finkelstein, "Laminar pipe flow with injection and suction through a porous wall," Trans. ASME: Ser. E **78**(3), 719–724 (1956).
[26] S. W. Yuan, "Further investigation of laminar flow in channels with porous walls," J. Appl. Phys. **27**(3), 267–269 (1956).
[27] J. R. Sellars, "Laminar flow in channels with porous walls at high suction Reynolds numbers," J. Appl. Phys. **26**(4), 489–490 (1955).
[28] R. M. Terrill and P. W. Thomas, "On laminar flow through a uniformly porous pipe," Appl. Sci. Res. **21**, 37–67 (1969).
[29] R. M. Terrill, "Laminar flow through a porous tube," J. Fluids Eng. **105**, 303–307 (1983).
[30] R. M. Terrill, "On some exponentially small terms arising in flow through a porous pipe," Q. J. Mech. Appl. Math. **26**(3), 347–354 (1973).
[31] J. Majdalani, "Vorticity dynamics in isobarically closed porous channels. Part I: standard perturbations," J. Propul. Power **17**(2), 355–362 (2001).
[32] J. Majdalani and T. S. Roh, "Vorticity dynamics in isobarically closed porous channels. Part II: space-reductive perturbations," J. Propul. Power **17**(2), 363–370 (2001).
[33] R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena* (J. Wiley, New York, 1960).
[34] J. Majdalani and G. A. Flandro, "The oscillatory pipe flow with arbitrary wall injection," Proc. R. Soc. **458**(2022), 1621–1651 (2002).
[35] R. H. Cantrell and R. W. Hart, "Interaction between sound and flow in acoustic cavities: mass, momentum, and energy considerations," J. Acoust. Soc. Am. **36**, 697–706 (1964).
[36] G. A. Flandro, "Effects of vorticity on rocket combustion stability," J. Propul. Power **11**(4), 607–625 (1995).
[37] J. Majdalani, "The oscillatory channel flow with arbitrary wall injection," J. Appl. Math. Phys. **52**(1), 33–61 (2001).
[38] J. Majdalani and T. S. Roh, "The oscillatory channel flow with large wall injection," Proc. R. Soc. **456**(1999), 1625–1657 (2000).
[39] J. Majdalani and W. K. Van Moorhem, "Improved time-dependent flowfield solution for solid rocket motors," AIAA J. **36**(2), 241–248 (1998).
[40] J. Majdalani, G. A. Flandro, and T. S. Roh, "Convergence of two flowfield models predicting a destabilizing agent in rocket combustion," J. Propul. Power **16**(3), 492–497 (2000).
[41] J. Majdalani and W. K. Van Moorhem, "Laminar cold-flow model for the internal gas dynamics of a slab rocket motor," J. Aerosp. Sci. Technol. **5**(3), 193–207 (2001).
[42] G. A. Flandro, "Effects of vorticity transport on axial acoustic waves in a solid propellant rocket chamber," in *Combustion Instabilities Driven by Thermo-Chemical Acoustic Sources* (American Society of Mechanical Engineers, New York, 1989), Vol. NCA 4, HTD 128, pp. 53–61.
[43] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[44] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, New York, 1964).

[45] T. S. Roh, I. S. Tseng, and V. Yang, ''Effects of acoustic oscillations on flame dynamics of homogeneous propellants in rocket motors,'' J. Propul. Power **11**(4), 640–650 (1995).
[46] N. Rott, ''Theory of time-dependent laminar flows,'' in *High Speed Aerodynamics and Jet Propulsion–Theory of Laminar Flows*, edited by F. K. Moore (Princeton U.P., Princeton, NJ, 1964), Vol. IV, pp. 395–438.
[47] H. Schlichting, *Boundary-Layer Theory*, 7th ed. (McGraw-Hill, New York, 1979).

Torsional wave experiments with a new magnetostrictive transducer configuration

Yoon Young Kim, Chan Il Park, Seung Hyun Cho, and Soon Woo Han
School of Mechanical and Aerospace Engineering and National Creative Research Initiatives Center for Multiscale Design, Seoul National University, Shinlim-Dong, San 56-1, Kwanak-Gu Seoul 151-742, Korea

(Received 20 April 2004; revised 13 March 2005; accepted 15 March 2005)

For the efficient long-range nondestructive structural health inspection of pipes, guided waves have become widely used. Among the various guided wave modes, the torsional wave is most preferred since its first branch is nondispersive. Our objective in this work is to develop a new magnetostrictive transducer configuration to transmit and receive torsional waves in cylindrical waveguides. The conventional magnetostrictive transducer for the generation and measurement of torsional waves consists of solenoid coils and a nickel strip bonded circumferentially to test pipes. The strip must be premagnetized by a permanent magnet before actual measurements. Because of the premagnetization, the transducer is not suitable for the long-term on-line monitoring of pipes buried underground. To avoid the cumbersome premagnetization and to improve the transduction efficiency, we propose a new transducer configuration using several pieces of nickel strips installed at 45° with respect to the pipe axis. If a static bias magnetic field is also applied, the transducer output can be substantially increased. Several experiments were conducted to study the performance of the proposed transducer configuration. The proposed transducer configuration was also applied for damage detection in an aluminum pipe.
[DOI: 10.1121/1.1904304]

PACS numbers: 43.20.Mv, 43.20.Gp, 43.35.Cg [AJZ] Pages: 3459–3468

I. INTRODUCTION

The guided-wave technology has received much attention recently as a powerful tool for the nondestructive inspection of cylindrical waveguides such as pipes and tubes.[1–6] Since guided waves can travel over several meters along the waveguide axis, the guided-wave technology is very efficient in inspecting a large portion of a waveguide. Several wave modes such as longitudinal, torsional, and flexural modes can be utilized for nondestructive inspection, but the torsional wave mode is preferred because its first branch is nondispersive and favorable for signal processing. Therefore, an efficient generation of the torsional wave is perhaps an important issue in the torsional wave-based nondestructive evaluation technology.

As far as the generation and measurement of the torsional wave are concerned, two approaches are available: one approach based on piezoelectric transducers[7] and the other based on magnetostrictive transducers.[8] Each approach has its own advantages and disadvantages, but we are concerned with the development of a new magnetostrictive transducer because magnetostrictive transducers are cost effective and easy to install. The magnetostrictive transducers use the coupling effect between the elastic deformation and the magnetic field of ferromagnetic materials. Although magnetostrictive transducers (including sensors) have been applied and studied in many cases,[9–14] their applications for torsional wave generation were made only recently by Kwun.[8]

The configuration of Kwun's transducer for generating torsional waves in a cylindrical waveguides is shown in Fig. 1(a). The transducer consists of a nickel strip and a solenoid coil surrounding the strip. The strip is bonded circumferentially to a test specimen such as a pipe and a permanent magnet is rubbed on the nickel strip for premagnetization. (Any material exhibiting strong magnetization can be used as the strip material, but nickel is easily available and cost effective.) The premagnetization will induce static magnetic field strength in the circumferential direction, which is indicated by H_S in Fig. 1(b). When alternating current is sent through the solenoid coil, the alternating magnetic field strength (H_D) is also developed on the nickel strip in the z axis direction. If the magnitudes of H_D and H_S are almost the same, the resulting magnetic strength vector will point in the direction that is about 45° from the z axis, as illustrated in Fig. 1(b).

By developing a new magnetostrictive transducer to generate torsional waves, Kwun[8] made a breakthrough in the field. Nevertheless, the configuration of his transducer has some drawbacks. First, the nickel strip always needs be premagnetized before the transducer is used. When the transducer is to be used for long-term on-line monitoring of underground pipes, the strip needs periodic magnetization, which is difficult to achieve. Second, if the magnitudes of H_D and H_S are not of the same order, undesirable wave modes are also generated in addition to the torsional wave mode.

To overcome the above-mentioned drawbacks, a new transducer configuration shown in Fig. 2 is proposed. Several pieces of nickel strips are attached to the test specimen with the alignment angle α equal to 45°. Since the relative permeability of the nickel strips is higher than that of the test specimen, most of the magnetic flux by the solenoid coil flows along nickel strips. Therefore, the elastic deformation of the nickel strips developed by the magnetostriction effect

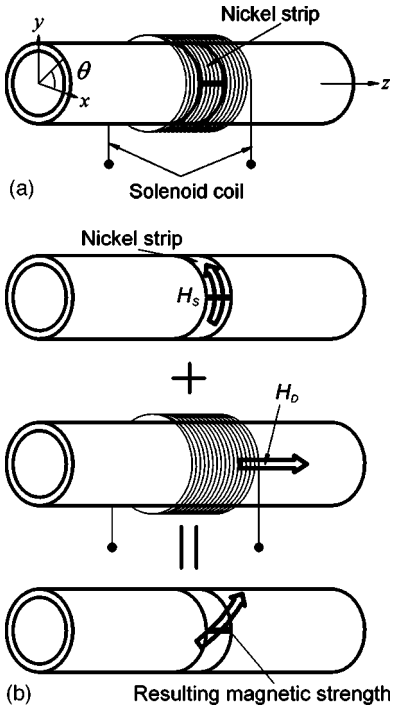


FIG. 1. Kwun's magnetostrictive transducer for generating and measuring torsional waves. (a) The schematic configuration; (b) the resulting direction of the magnetic field strength as the sum of H_S and H_D.

results in the main elastic deformation of the test specimen in the direction parallel to the strip alignment direction. Obviously, the generated strain will develop torsional waves along the specimen axis.

The main difference between Kwun's transducer and the proposed transducer is simply the alignment angle of the nickel strip. However, the alignment change has a significant impact on the transducer characteristics; premagnetization is not needed and the generated wave mode is insensitive to the magnitude of the current input to the solenoid coil. The proposed transducer can perform even without the bias solenoid, although the applied bias field will definitely improve the transducer performance. The effect of the bias magnetic field will be investigated in the subsequent section.

Earlier, Ohzeki and Mashine[15] used skew-oriented ferromagnetic patches attached to the test specimen with respect to the test specimen axis. Their motivation was to estimate the torque transmitted in the shaft of a milling machine. The proposed transducer in this work, however, can not only measure but also generate torsional waves. Damage location estimations in aluminum pipes will be also considered as a typical application problem. To verify the performance of the

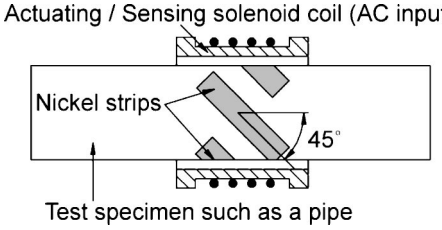


FIG. 2. The proposed magnetostrictive transducer for generating and measuring torsional waves.

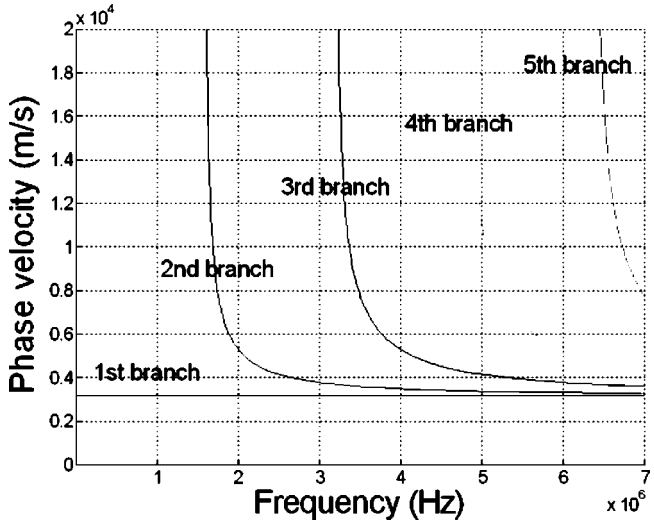


FIG. 3. The phase-velocity dispersion curve for the axisymmetric torsional wave in an cylindrical aluminum shell with the inner and outer radii $a=11.5$ mm and $b=12.5$ mm.

proposed transducer, several experiments were performed.

II. THEORETICAL BACKGROUND

In this section, the theory on the guided torsional waves and the magnetostrictive effects will be discussed.

A. Guided torsional waves

Thin-walled pipes are used as waveguides throughout this investigation, so the mechanics of elastic waves as well as their dispersion characteristics will be given mainly in cylindrical shells shown in Fig. 1(a). Figure 3 shows the dispersion curve as the $\nu_p-\omega$ relation (ν_p: the phase velocity, ω: angular frequency). As can be seen from Fig. 3, the first branch, the lowest-energy branch, is nondispersive. Therefore, the phase velocity of the wave belonging to the first branch is independent of frequency. If the excitation frequency is not much higher than the first cutoff frequency (1.6 MHz in this case), the first branch becomes the main carrier of wave energy. Therefore, the excitation pulse shape will be nearly preserved. Since no other wave mode or branch has the nondispersive characteristics, it is best to use the pulse that can be decomposed within the first branch of the torsional wave mode for long-range damage detection. For general discussions on the dispersion characteristics, see Achenbach,[16] Miklowitz,[17] Graff,[18] or Rose.[19]

B. Magnetostrictive effects

The proposed transducer uses the magnetostrictive effects in actuating and measuring torsional waves, so the physics of the magnetostrictive effects such as the Joule effect and the Villari effect should be explained. The Joule effect[20] refers to the phenomenon of the dimension change of a piece of ferromagnetic material when it is placed under a magnetic field. The Villari effect[21] represents the inverse phenomenon of the Joule effect.

The Joule effect and the Villari effect may be expressed by the following two equations for one-dimensional situations:

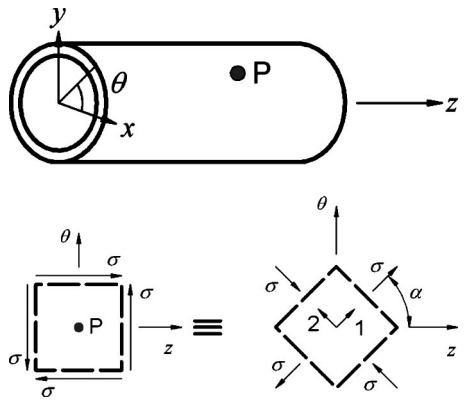


FIG. 4. The stress state corresponding to pure torsion in a pipe. (a) The generic point P on the pipe surface; (b) the two-dimensional view of the stress state.

$$\varepsilon = \frac{\sigma}{E^H} + q^* H, \tag{1}$$

$$B = \mu^\sigma H + q\sigma, \tag{2}$$

where ε, σ, B, and H represent the strain, stress, magnetic flux density, and magnetic strength, respectively. The material constants E^H, q^*, μ^σ, and q denote the Young's modulus under a constant magnetic strength, the coupling coefficient of the Joule effect, the permeability under a constant stress and the coupling coefficient of the Villari effect, respectively. A general theory on the magnetostrictive effect including explanations on hysteresis and irreversibility can be found in Jiles.[22] The elastic wave or deformation in a ferromagnetic material can be easily converted to the voltage change of the solenoid surrounding the material.[12]

III. PROPOSED MAGNETOSTRICTIVE TRANSDUCER FOR TORSIONAL WAVES

A. Transducer configuration and first-order stress analysis

To explain the mechanism to generate torsional waves by the Joule effect in a pipe, consider a generic point P on the pipe surface illustrated in Fig. 4(a). When the pipe shown in Fig. 4(a) is at the state of pure torsion, the stress state at P in the z-θ plane should look like the state shown in Fig. 4(b). The pure shear state in the z-θ coordinate system can be represented by two normal stresses $(\sigma, -\sigma)$ in the principle axes 1 and 2 that are oriented $\pm 45°$ from the z axis. Therefore, if the normal stresses having the opposite signs are applied along the two principle axes, pure torsional waves can be generated along the pipe axis.

Based on the simple stress analysis in Fig. 4(b), we propose to align the ferromagnetic strip (made of nickel) at 45° from the pipe axis. Figure 5(a) shows four nickel strips bonded on the outer surface of an aluminum pipe, and Fig. 5(b) shows the magnetostrictive transducer consisting of the nickel strips and the solenoid coil. The solenoid coil serves both as the actuating coil and the sensing coil.

The aluminum pipe will be excited by the elastic deformation of the nickel strip through the magnetostrictive effect. Therefore, it is important to investigate the direction of the

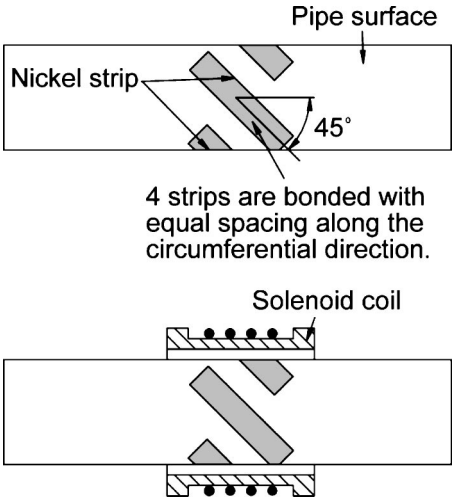


FIG. 5. Nickel strips bonded on the surface of pipe (a) with and (b) without the actuating and sensing solenoid coil installed.

flux line in the strip when elastic current flows through the solenoid coil. Figure 6 shows the pattern of the flux line in the z-θ plane. For the analysis, a two-dimensional linear static model was used. The actual numerical calculation was performed by ANSYS.[23] In Fig. 6, the magnetic flux in the strip flows mainly along axis 1, since the permeability of nickel is much larger than those of aluminum and air. Therefore, the Joule effect causes the strip to deform mainly along the axis 1 direction when electric current flows through the solenoid coil.

To investigate the wave generated by the nickel strip, let us consider the elastic deformation of the nickel strip at the moment of the electric current input to the solenoid coil and carry out a first-order stress analysis. The uniform stretch (or contraction) along axis 1 and the accompanied uniform contraction (or stretch) along axis 2 developed in the nickel strip before being bonded may be expressed approximately by the following strain components:

$$\varepsilon_{11} = qH, \quad \varepsilon_{22} = -\nu_N \varepsilon_{11} = -\nu_N qH, \quad \varepsilon_{12} = 0, \tag{3}$$

where ν_N is Poisson's ratio of nickel.

When the nickel strip is bonded to the aluminum pipe, the elastic deformation of the nickel strip develops stress in

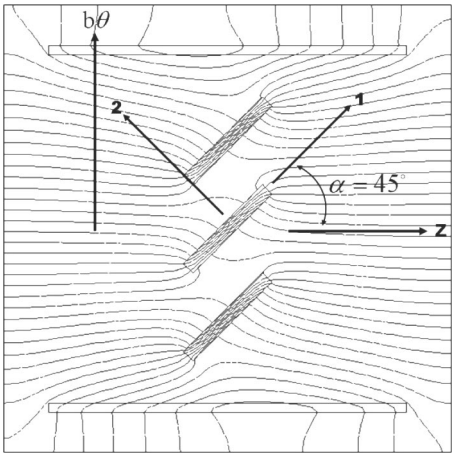


FIG. 6. The magnetic flux lines.

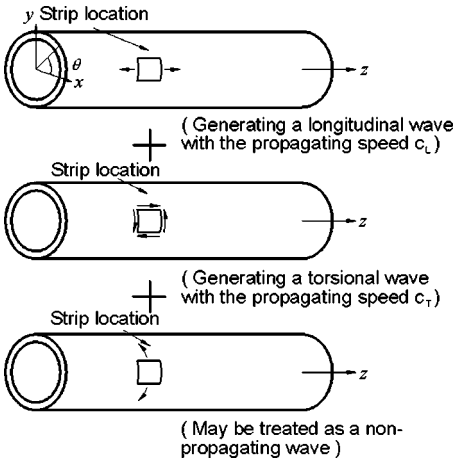


FIG. 7. The decomposition of the stress state in a pipe developed by the nickel strip deformation.

the aluminum pipe. The evaluation of the magnitude of the stress requires a somewhat complicated analysis, so some assumptions will be made for the first-order analysis. If $l_N \approx n w_N$ (l_N, w_N: length and width of each strip, n: number of the nickel strip) in Fig. 2, the stress developed in the pipe where the nickel strip is attached may be assumed as

$$\sigma_{11} \approx \sigma_0, \quad \sigma_{22} \approx -\nu_N \sigma_0, \quad \sigma_{12} \approx 0, \tag{4}$$

where σ_0 stands for the normal stress component in axis 1.

To convert the stress components expressed along axes 1 and 2 into those along the z-θ coordinate system, the following transformation formula will be used (see, e.g., Timoshenko and Goodier[24]):

$$\begin{aligned}
\sigma_{i'i'} &= \frac{1}{2}(\sigma_{ii}+\sigma_{jj}) + \frac{1}{2}(\sigma_{ii}-\sigma_{jj})\cos 2\alpha + \sigma_{ij}\sin 2\alpha \\
\sigma_{j'j'} &= \frac{1}{2}(\sigma_{ii}+\sigma_{jj}) - \frac{1}{2}(\sigma_{ii}-\sigma_{jj})\cos 2\alpha - \sigma_{ij}\sin 2\alpha \\
\sigma_{i'j'} &= -\frac{1}{2}(\sigma_{ii}-\sigma_{jj})\sin 2\alpha + \sigma_{ij}\cos 2\alpha,
\end{aligned} \tag{5}$$

where α is the angle between the two orthogonal coordinate systems $[(i,j),(i',j')]$. Substituting $(i',j')=(z,\theta)$, $(i,j)=(1,2)$ and $\alpha = 45°$, the stress components in Eq. (4) are written as

$$\begin{aligned}
\sigma_{zz} &= \frac{1-\nu_N}{2}\sigma_0, \\
\sigma_{\theta\theta} &= \frac{1-\nu_N}{2}\sigma_0, \\
\sigma_{z\theta} &= -\frac{1+\nu_N}{2}\sigma_0.
\end{aligned} \tag{6}$$

Though the pure shear state ($\sigma_{zz}=\sigma_{\theta\theta}=0$, $\sigma_{z\theta}\neq 0$) is desirable, the normal stress is also developed, as can be seen in Eq. (6). Therefore, not only the torsional wave (due to $\sigma_{z\theta}$), but also the longitudinal wave (due to σ_{zz}) will be generated. Figure 7 shows the stress state decomposed into three stress components.

To study the stress wave propagation, it is very important to note that the longitudinal wave speed $c_L = \sqrt{E_A/\rho_A}$ is different from the torsional wave speed $c_T = \sqrt{E_A/2\rho_A(1+\nu_A)}$ (E_A, ρ_A, ν_A: Young's modulus, density, Poisson's ratio of aluminum). Therefore, σ_{zz} component generating the longitudinal wave and $\sigma_{z\theta}$ component generating the torsional wave separate as they propagate. Since $c_L > c_T$, the longitudinal wave travels ahead of the torsional wave.

To continue the first-order stress analysis, one may neglect the dynamic effect of the propagating waves, and the damping effect occurring during wave propagation. When the generated longitudinal stress wave is measured by the transducer, the stress state will be approximated as

$$\sigma_{zz} = \frac{1-\nu_N}{2}\sigma_0, \quad \sigma_{\theta\theta}=0, \quad \sigma_{z\theta}=0. \tag{7}$$

Then the following stress component will be measured through the nickel strip:

$$\sigma_{11}^L = \frac{1-\nu_N}{4}\sigma_0, \tag{8}$$

where the superscript L stands for the longitudinal wave. In obtaining the result in Eq. (8), Eq. (5), and Eq. (6) are used, with $(i',j')=(1,2)$, $(i,j)=(z,\theta)$, $\alpha=-45°$. Likewise, the magnitude of the σ_{11}^T stress component due to the torsional wave can be calculated as

$$\sigma_{11}^T = \frac{1+\nu_N}{2}\sigma_0. \tag{9}$$

From the above stress analysis, the following conclusions can be drawn.

(1) By the proposed transducer configuration, both the torsional wave and the longitudinal wave are simultaneously generated. However, the magnitude of the torsional stress σ_{11}^T is much larger than that of the longitudinal stress σ_{11}^L. The relative ratio, with dynamic and damping effects neglected, is approximated as

$$\left|\frac{\sigma_{11}^T}{\sigma_{11}^L}\right|_{FOA} = \left|\frac{\frac{1+\nu_N}{2}\sigma_0}{\frac{1-\nu_N}{4}\sigma_0}\right| = \frac{2(1+\nu_N)}{1-\nu_N}$$

$$\approx 3.80 \quad (\text{for} \quad \nu_N = 0.31). \tag{10}$$

In Eq. (10), the subscript "FOA" is used to emphasize that the result is predicted by the first-order analysis.

(2) The torsional wave will be preceded by the longitudinal wave of a small magnitude due to the difference between their wave speeds. So, the torsional wave can be distinguished against the longitudinal wave.

Though the stress analysis is based on some assumptions, the present analysis shows the characteristics of the waves measured by the proposed transducer.

The photo of the proposed transducer is shown in Fig. 8. The bias coil shown in Fig. 8(b) is installed to supply the static bias magnetic field. The bias field is very useful to increase the transducer output, as shall be shown later.

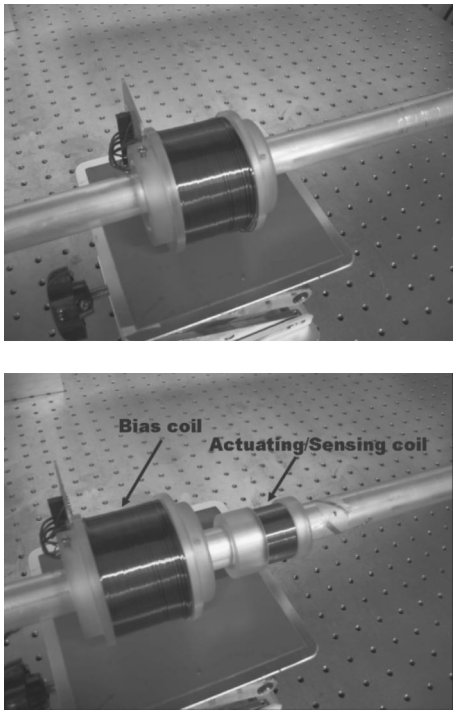


FIG. 8. The photo of the proposed magnetostrictive transducer. (a) At the assembled state; (b) at the disassembled state.

1. Comparison with Kwun's transducer

When Kwun's transducer illustrated in Fig. 1 is used, the strain developed in the unbounded nickel strip may be approximated as

$$\varepsilon_{11}=q\sqrt{H_S^2+H_D^2},\quad \varepsilon_{22}=-\nu_N\varepsilon_{11},$$

$$\varepsilon_{12}=0,\quad \text{and}\quad \tan\alpha=\frac{H_S}{H_D}. \tag{11}$$

In this case, the directions of principle axes 1 and 2 depend on the magnitude of H_S/H_D. Unless H_D is adjusted to be the same as H_S, i.e., $\alpha=45°$, torsion-dominant waves cannot be generated. If H_D is much larger than H_S, for instance, the generated wave by the Kwun's transducer will be dominated by the longitudinal wave mode. Since the H_S field is created by rubbing a permanent magnet on the nickel strip along the circumferential direction, it is difficult to quantify the magnitude of H_S. Furthermore, nickel is usually demagnetized in a relatively short period, so the permanent magnet rubbing technique is not desirable for the long-term on-line inspection especially in pipes buried underground.

B. Experimental setup

Figure 9 shows the schematic diagram of the experimental setup. The transducer in Fig. 9 is the proposed transducer shown in Fig. 8. The nickel strips are bonded to the aluminum pipe by epoxy (Model: 3M DP460). The dimension of the pipe specimen and the location of the proposed transducer are also shown in Fig. 9. The selected specification of the proposed transducer is given in Fig. 10. Better transducer performance obtained with other specifications will be investigated thoroughly in future work.

IV. EXPERIMENTAL STUDY

In this section, the performance and the characteristics of the proposed transducer will be investigated. In the existing magnetostrictive transducer applications, the following sine waveform was used:

$$f_{\mathrm{sp}}(t)=\begin{cases}\sin\eta t, & 0\leqslant t\leqslant\dfrac{2\pi}{\eta}m,\\ 0, & \text{elsewhere,}\end{cases} \tag{12}$$

where t is time and η is the center frequency and m usually takes on the value of 1 or 2. When $m=1$, a single sine pulse is generated. As pointed out by Hong and Kim,[25] the energy of $f_{\mathrm{sp}}(t)$ is not sufficiently concentrated on the center frequency η; $f_{\mathrm{sp}}(t)$ is thus not efficient for accurate damage location estimation. To concentrate the energy of the generated pulse at a given center frequency η, we employ the Gabor pulse $f_{\mathrm{Gp}}(t)$:

$$f_{\mathrm{Gp}}(t)=\frac{1}{(\pi\hat{\sigma})^{1/4}}e^{-t^2/2\hat{\sigma}^2}\cos\eta t. \tag{13}$$

The parameter $\hat{\sigma}$ controls the spread of the pulse in time. The shape of the Gabor pulse for $\hat{\sigma}=0.000\,083$ and $\eta=60$ kHz is shown in Fig. 11(a), and the actual waveform coming out of the power amplifier is plotted in Fig. 11(b). Figure 11(b) shows that the Gabor-type pulse generated for experiments is almost the same as the theoretical pulse.

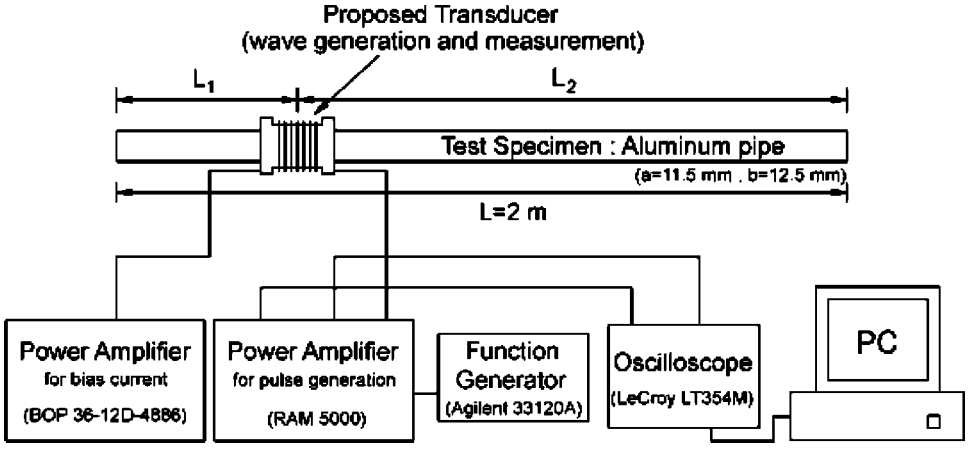


FIG. 9. The schematic diagram of the experimental setup.

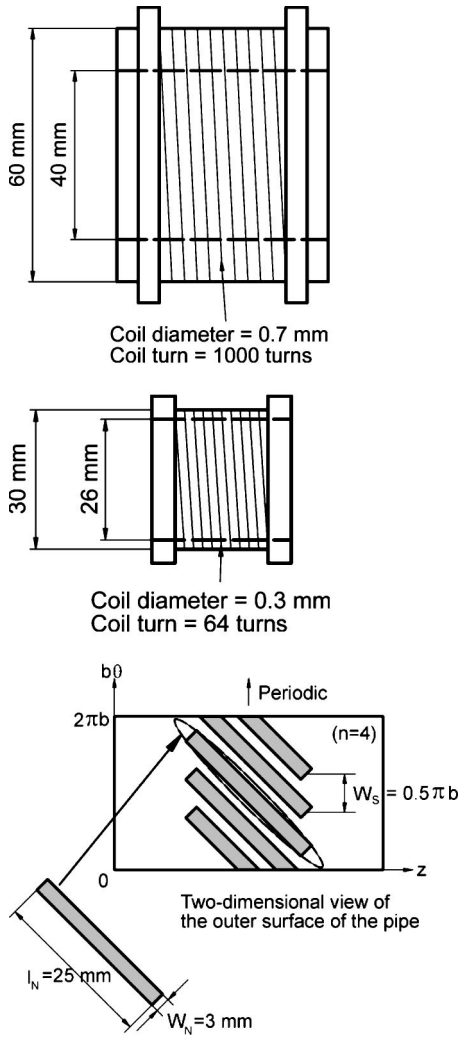


FIG. 10. The specification of the proposed transducer. (a) Bias coil; (b) actuating/sensing coil; and (c) nickel strip (thickness=0.15 mm).

A. Wave measurement and comparison with Kwun's transducer

In characterizing the measured pulses, the following notations will be used:

s^-: pulse traveling to the left at the initial state

s^+: pulse traveling to the right at the initial state

d in $s^-[d]$ or $s^+[d]$: the distance the pulse s^- or s^+ travels

Symbols T and L of the symbol s: torsional mode and longitudinal mode, respectively.

For instance, $s^-[2L_1]$ is the pulse initially traveling to the left, which arrives at $t=2L_1/c_T$ after it travels a distance of $2L_1$, as illustrated in Fig. 12(a). At time $t=2L/c_T$, two pulses s^+ and s^- will merge, so the magnitude of the signal become large.

A signal captured by the proposed transducer for the Gabor pulse with $\eta=60$ kHz is shown in Fig. 13(a). Some pulses of the signal in Fig. 13(a) are zoomed in Fig. 13(b). Unless stated otherwise, the following numerical data are used:

i_B (DC input to the bias solenoid) $=0.5$ A

$\hat{\sigma}=0.000\,083$

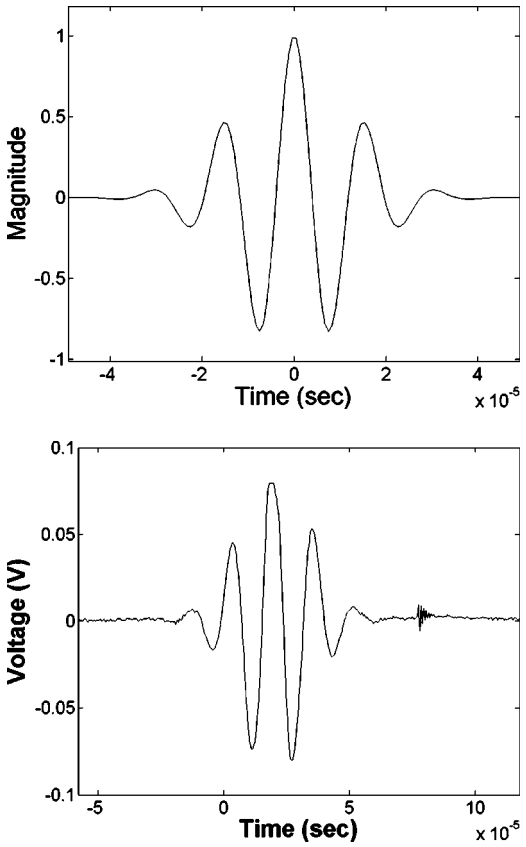


FIG. 11. The Gabor waveform. (a) The theoretical signal; (b) the actual signal coming from the power amplifier.

The short-time Fourier transform of the measured signal in Fig. 13(a) is presented by Fig. 13(c). Figure 13(c) clearly shows that the energy of each pulse is centered at the excitation frequency $\eta=60$ kHz. The nondispersive characteristic of the first branch of the torsional wave is revealed by the straight vertical lines in the time–frequency plane, around at $t=2L_1/c_T=0.316\times10^{-6}$ s, $t=2L/c_T=1.254\times10^{-6}$ s, etc.

As explained in Sec. III A, the signal shown in Fig. 13(a) contains the pulses corresponding to the longitudinal wave mode. These pulses, marked by s_L in Fig. 13(a), may not be easily identified because their magnitudes are small. As longitudinal waves are dispersive, they spread out as they propagate. Thus, it may be appropriate to choose the magni-

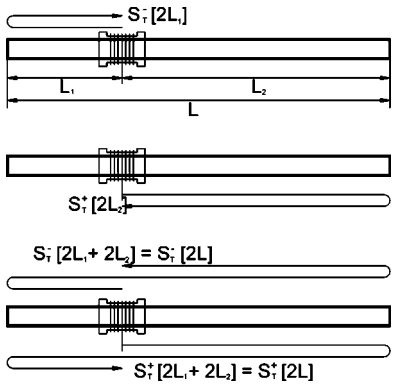


FIG. 12. The illustration of the traveling histories of pulses arriving at (a) $t=2L_1/c_T$; (b) $t=2L_2/c_T$; (c) $t=2L/c_T$.

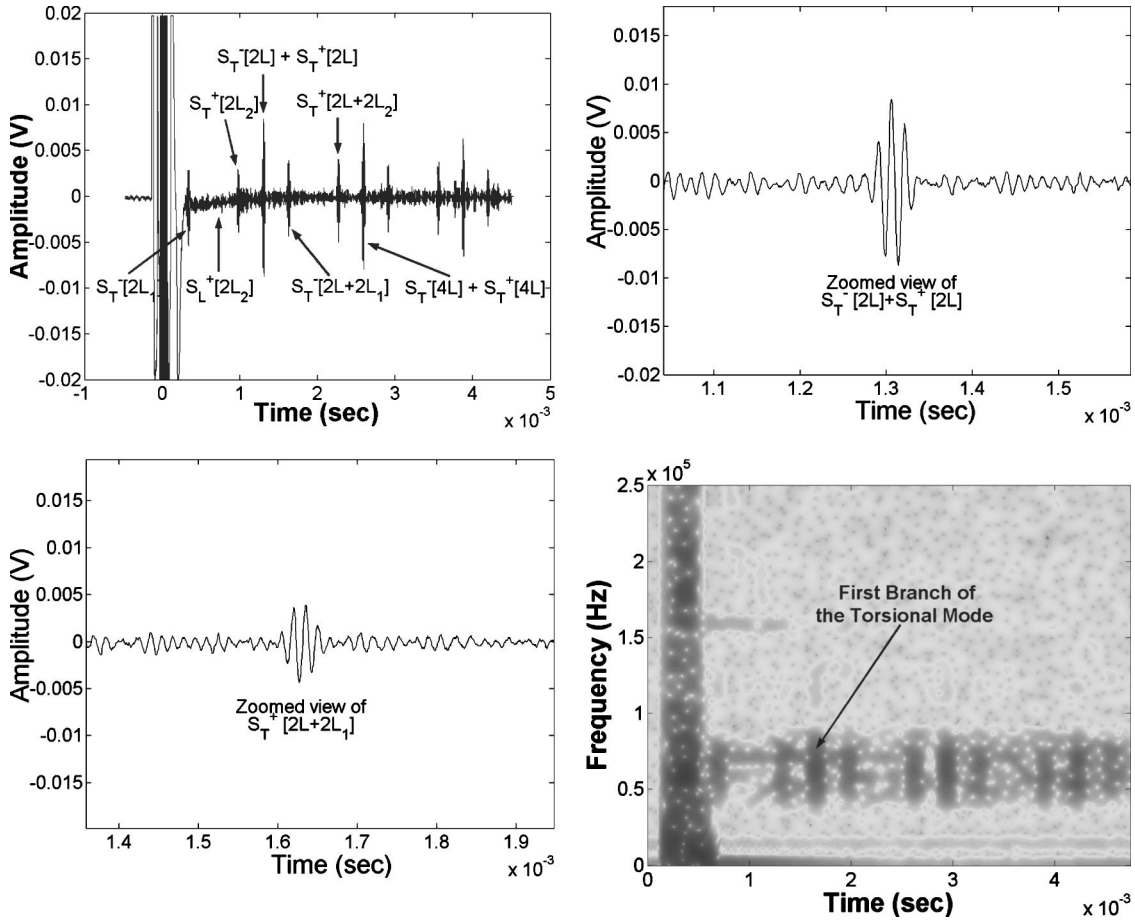


FIG. 13. The measured signal by the proposed transducer. (The center frequency η of the Gabor-type pulse: 60 kHz). (a) The time history; (b) the zoomed view of $s^-[2L+2L_1]$ and $s^-[2L]+s^+[2L]$; (c) the spectrogram of the signal in (a) by the short-time Fourier transform.

tude of the largest peak of s_L as the longitudinal wave level. This peak value will be denoted by σ_{11}^L.

It is important to compare σ_{11}^T, the peak torsional wave magnitude with σ_{11}^L in checking the transducer performance. To this end, the signal in Fig. 13(a) is zoomed, and the arrival time of the longitudinal wave mode is predicted by using the longitudinal wave speed c_L. (The longitudinal wave pulses may be seen better at different excitation frequencies: see Fig. 15, later.) The estimated ratio of $|\sigma_{11}^T/\sigma_{11}^L|$ from the experimentally measured signal in Fig. 13(a) is $|\sigma_{11}^T/\sigma_{11}^L|_{\text{exp}}=3.6$, which is not far from $|\sigma_{11}^T/\sigma_{11}^L|_{\text{FOA}}=3.80$.

To compare the performance of the proposed transducer and that of Kwun's transducer, the same wave measurement experiments were conducted with Kwun's transducer shown in Fig. 1(a). For quantitative comparison, the same pipe and actuating/sensing coil are used. Furthermore, the amount of nickel for Kwun's strip is the same as that for the proposed transducer. Since Kwun's strip is bonded circumferentially around the outer pipe surface, the strip width w_N^{Kwun} for Kwun's transducer is chosen as

$$w_N^{\text{Kwun}}=4\times l_N\times w_N/2\pi b. \tag{14}$$

Note that four pieces of the nickel strip are used in the proposed transducer.

The overall characteristics of the pulses measured by Kwun's in Fig. 14 are the same as those in Fig. 13(a) and Fig. 13(c). However, the magnitude of the measured pulse by Kwun's transducer is quite smaller than that by the proposed transducer. The peak-to-peak magnitudes ($V_{p\text{-}p}$) of the pulse $s^-[2L]+s^+[2L]$, for instance, are as follows:

By the proposed transducer: $V_{p\text{-}p}=0.017\,18$ V;
by Kwun's transducer: $V_{p\text{-}p}=0.004\,21$ V.

One reason why the proposed transducer gives larger peak-to-peak values than Kwun's transducer may be due to the applied static bias magnet field. In the next section, we will investigate the effect of the bias magnetic field on the output of the proposed transducer.

B. The effects of the static bias magnetic field and the excitation frequency

The static bias magnetic field affects the transducer performance. To investigate the effect of the bias magnetic field, the input electric current i_B to the bias solenoid coil is varied. Nine values of i_B are considered and the peak-to-peak values ($V_{p\text{-}p}$) of s^-+s^+ at $t=2L/c_T$ are tabulated in Table I. In adjusting the magnitude of i_B, the magnetostrictive effect of nickel relating the magnetic field strength H to the strain ε is considered.[26] As shown by Table I, the bias magnetic field significantly affects the transducer output. The peak-to-peak value monotonically increases for small values of i_B and

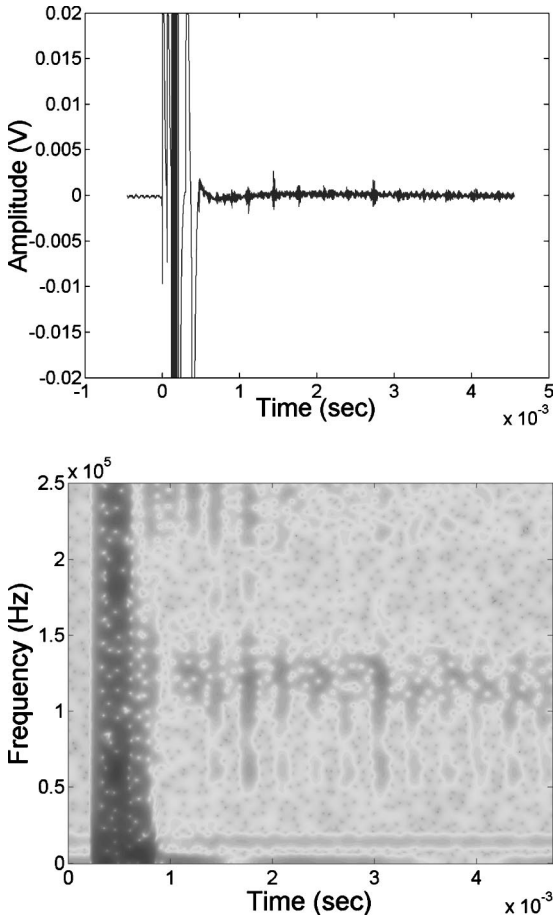


FIG. 14. The measured signal by Kwun's transducer ($\eta = 60$ kHz). (a) The time history; (b) the spectrogram by the short-time Fourier transform.

remains almost the same after $i_B \approx 0.5$ A. Based on this finding, $i_B = 0.5$ A was used for the present investigation.

One advantage of the proposed transducer is apparent from this experiment. Unlike Kwun's configuration, the bias magnetic field is easily controlled by the change of i_B. Therefore, the transducer performance can be always maximized. Obviously, the bias current maximizing the transducer output will be different for different excitation frequencies, but this experiment will suffice to show the importance of the appropriate bias magnetic field and to demonstrate the versatility of the proposed transducer configuration.

TABLE I. The peak-to-peak values $V_{p\text{-}p}$ of $s^- + s^+$ at $t = 2L/c_T$ for varying values of i_B.

i_B (A)	Estimated H_B (kA/m)	$V_{p\text{-}p}$ (V)
0	0	0.0025
0.1	2	0.0076
0.2	4	0.0094
0.3	6	0.0106
0.4	8	0.0107
0.5	10	0.0108
0.6	12	0.0101
0.7	14	0.0100
0.8	16	0.0094

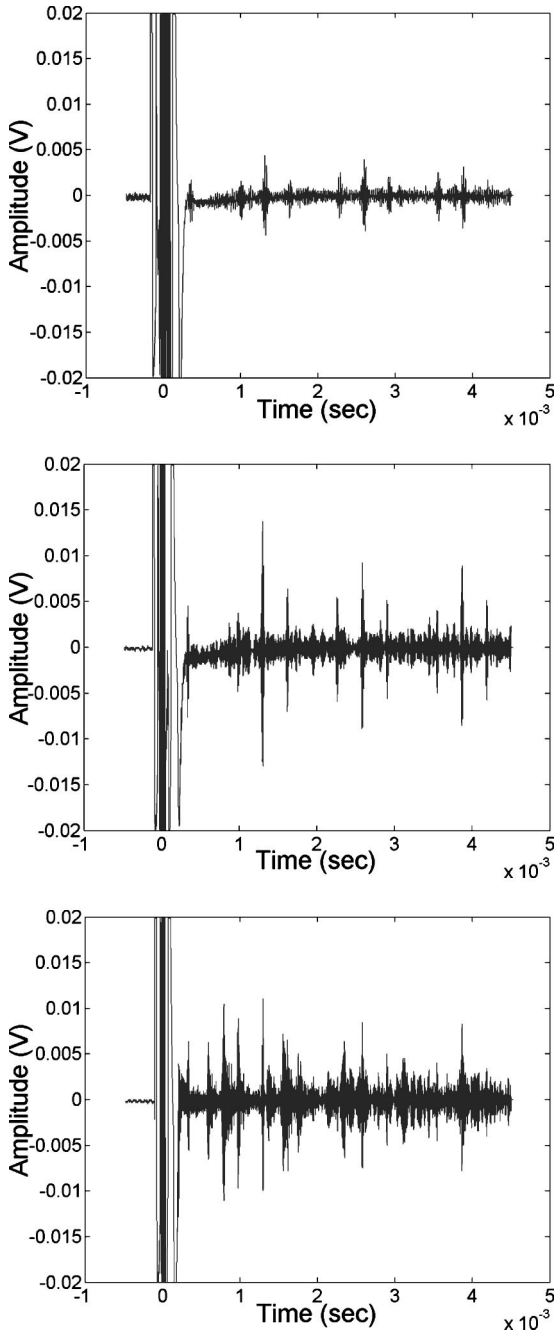


FIG. 15. The measured signals at different excitation frequencies. (a) $\eta = 40$ kHz; (b) 80 kHz; and (c) 120 kHz.

Now the effect of the excitation frequency on the measured signal will be investigated. The measured signals for $\eta = 40$ kHz, 80 kHz, and 120 kHz are shown in Fig. 15. The bias current was set at 0.5 A for all experiments. Figure 16 shows the variation of the peak-to-peak value of the torsional wave pulse. It also shows the peak-to-peak ratio between the torsional pulse and the longitudinal wave pulse.

The maximum peak-to-peak value of the torsional wave pulse was around $\eta \approx 95$ kHz, but the relative magnitude of the peak-to-peak value of the torsional wave pulse to the peak-to-peak value of the longitudinal wave pulse reaches its maximum at around $\eta \approx 60$ kHz. Therefore, some tradeoff between the two factors must be made in choosing the exci-

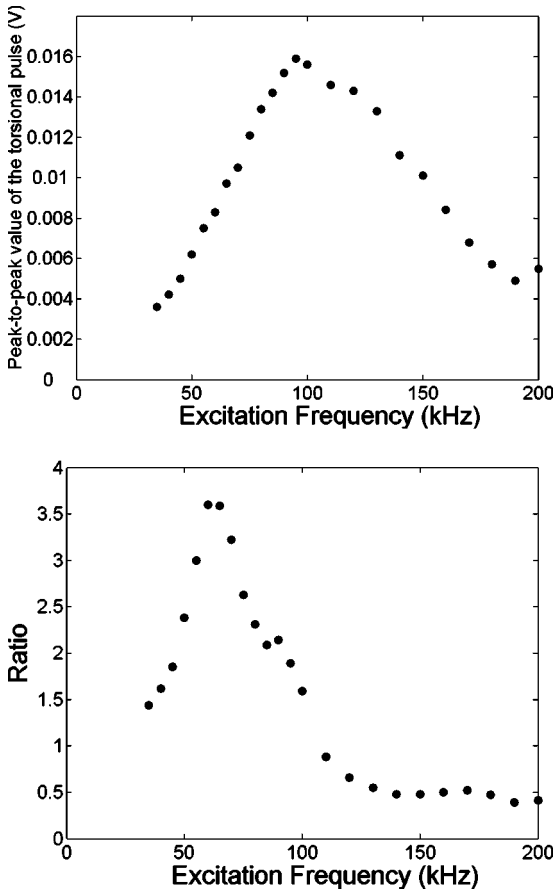


FIG. 16. The effect of the excitation frequency η on the transducer output. (a) The peak-to-peak value ($V_{p\text{-}p}$) of the torsional pulse; (b) the ratio of the peak-to-peak value of the torsional wave pulse to that of the longitudinal wave pulse.

tation frequency, depending on the actual field situation. However, this experiment shows the importance of the excitation frequency in maximizing the transducer performance.

To check the applicability of the proposed transducer in pipes having larger dimensions, a few experiments were conducted. The time signals obtained for pipes with ($d_0 = 50$ mm, $t = 1$ mm, $\eta = 25$ kHz) and ($d_0 = 50$ mm, $t = 3$ mm, $\eta = 25$ kHz) are plotted in Fig. 17. As illustrated in Fig. 17, the proposed transducer can be used for larger, thicker pipes. However, more experiments should be performed to apply the transducer to industrial problems involving thick-walled pipes (say, $t = 10$ mm).

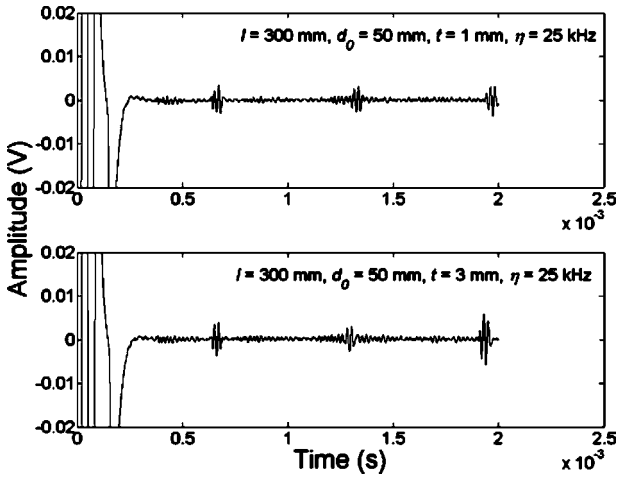


FIG. 17. The measured signals in pipes having ($d_0 = 50$ mm, $t = 1$ mm) and ($d_0 = 50$ mm, $t = 3$ mm). The center frequency η was selected so as to yield high T/L ratio and outputs.

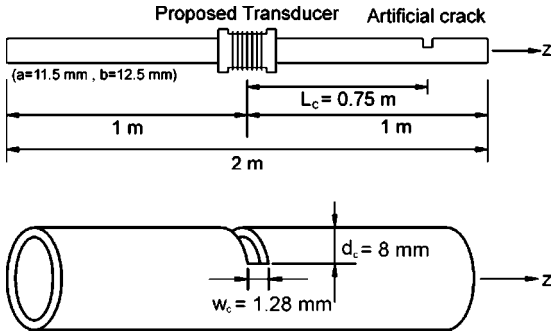


FIG. 18. The experimental setup for a damage detection application. (a) A cracked aluminum pipe; (b) crack dimension.

C. Damage detection application

As an application of the proposed transducer, damage detection in a pipe is considered. The experimental setup for damage detection is shown in Fig. 18(a) and the dimensions of an artificial crack are shown in Fig. 18(b). The specifications of the test pipe and the transducer are the same as those used in earlier experiments.

Figure 19 shows the measured signal by the proposed transducer, which contains the distinctive pulses reflected from the crack. Using the torsional wave speed $c_T = 3188$ m/s, the distance L_c from the transducer to the crack can be estimated. The result is summarized in Table II; the damage location was accurately determined from the signal measured by the proposed transducer. Although many other cracking situations should be investigated, this subject will be studied in the future. Nevertheless, the present damage detection problem suffices to demonstrate the effectiveness of the proposed transducer. (See Fig. 19).

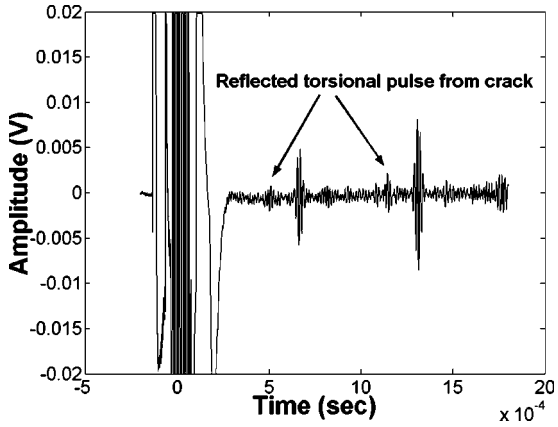


FIG. 19. The measured signal in a cracked aluminum pipe. The Gabor pulse with $\eta = 60$ kHz is used.

TABLE II. The estimation of the crack location (L_c) by the proposed transducer.

Exact	Estimated by experiment	Error
75 cm	76 cm	1.34%

V. CONCLUSIONS

The generation and measurement of torsional waves in cylindrical waveguides by a new magnetostrictive transducer was studied. The proposed transducer consisted of a solenoid coil and several nickel strips bonded on the test specimen at an alignment angle of 45°. A bias solenoid coil supplying the static magnetic field to the nickel strip can be installed to increase the transducer output. In comparison with the currently available transducer, the proposed one avoided cumbersome premagnetization of the nickel strip and produced bigger torsional signals. A first-order stress analysis was carried out to estimate the relative magnitude of the generated torsional stress wave compared to the magnitude of the generated longitudinal wave. The experiments conducted with the proposed transducer showed that the center frequency of the excitation pulse and the magnitude of the applied direct current input to the bias solenoid coil significantly affected the transducer performance. Finally, the proposed transducer was applied successfully to the damage detection in a long cylindrical pipe.

[1] M. G. Silk and K. F. Bainton, ''The propagation in metal tubing of ultrasonic wave modes equivalent to Lamb waves,'' Ultrasonics **17**, 11–19 (1979).
[2] J. L. Rose, Y. Cho, and J. J. Ditri, ''Cylindrical guided wave leakage due to liquid loading,'' in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti, 1994, pp. 259–266.
[3] J. L. Rose, J. J. Ditri, A. Pilarski, K. Rajana, and F. T. Carr, ''A guided wave inspection technique for nuclear stream generator tubing,'' NDT & E Int. **27**, 307–330 (1994).
[4] P. Cawley and D. Alleyne, ''The use of Lamb waves for the long range inspection of large structures,'' Ultrasonics **34**, 287–290 (1996).
[5] M. J. S. Lowe, D. N. Alleyne, and P. Cawley, ''Defect detection in pipes using guided waves,'' Ultrasonics **36**, 147–154 (1998).
[6] C. Aristégui, M. J. S. Lowe, and P. Cawley, ''Guided waves in fluid-filled pipes surrounded by different fluids,'' Ultrasonics **39**, 367–375 (2001).
[7] J. O. Kim and O. S. Kwon, ''Vibration characteristics of piezoelectric torsional transducers,'' J. Sound Vib. **264**, 453–473 (2003).
[8] H. Kwun, ''Method and apparatus generating and detecting torsional wave inspection of pipes or tubes,'' U.S. Patent 6429650 B1, 2002.
[9] H. Kwun and C. M. Teller, ''Magnetostrictive generation and detection of longitudinal, torsional, and flexural waves in a steel rod,'' J. Acoust. Soc. Am. **96**, 1202–1204 (1994).
[10] H. Kwun and K. A. Bartels, ''Experimental observation of wave dispersion in cylindrical shells via time-frequency analysis,'' J. Acoust. Soc. Am. **97**, 3905–3907 (1995).
[11] H. Kwun and K. A. Bartels, ''Magnetostrictive sensor technology and its applications,'' Ultrasonics **36**, 171–178 (1998).
[12] H. C. Lee and Y. Y. Kim, ''Wave selection using a magnetomechanical sensor in a solid cylinder,'' J. Acoust. Soc. Am. **112**, 953–960 (2002).
[13] S. H. Cho, Y. Kim, and Y. Y. Kim, ''The optimal design and experimental verification of the bias magnet configuration of a magnetostrictive sensor for bending wave measurement,'' Sens. Actuators, A **107**, 225–232 (2003).
[14] Y. Y. Kim, S. H. Cho, and H. C. Lee, ''Application of magnetomechanical sensors for modal testing,'' J. Sound Vib. **268**, 799–808 (2003).
[15] H. Ohzeki, A. Mashine, H. Aoyama, and I. Inasaki, ''Development of a magnetostrictive torque sensor for milling process monitoring,'' J. Manuf. Sci. Eng. **121**, 615–622 (1999).
[16] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1975), pp. 241–242.
[17] J. Miklowitz, *Elastic Waves and Waveguides* (North-Holland, Amsterdam, 1978), pp. 219–220.
[18] K. F. Graff, *Wave Motion in Elastic Solids* (Ohio State University Press, Columbus, 1975), pp. 475–480.
[19] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, New York, 1999), pp. 154–175.
[20] J. P. Joule, ''On the effect of magnetism upon the dimensions of iron and steel bars,'' Philos. Mag. **30**, 76 (1847).
[21] E. Villari, ''Change of magnetization by tension and by electric current,'' Ann. Phys. Chem. **126**, 87–122 (1865).
[22] D. C. Jiles, ''Theory of the magnetomechanical effect,'' J. Phys. D **28**, 1537–1546 (1995).
[23] ANSYS, ANSYS Lower-frequency Electromagnetic Analysis Guide, ANSYS, Inc., Canonsburg, PA, 2003.
[24] S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 3rd ed. (McGraw-Hill, New York, 1970), pp. 8–11.
[25] J. C. Hong, K. H. Sun, and Y. Y. Kim, ''The matching pursuit approach based on the modulated-Gaussian pulse for efficient guided-wave damage inspection,'' Smart Mater. Struct. **14**, 1–13 (2005).
[26] R. M. Borzorth, *Ferromagnetism* (IEEE Press, Piscataway, NJ, 1993), pp. 659–663.

Estimation of dynamic elastic constants from the amplitude and velocity of Rayleigh waves

Ana Bayón
Departamento de Física Aplicada a los Recursos Naturales, Universidad Politécnica de Madrid, E.T.S. Ingenieros de Minas, C./Ríos Rosas, 21, 28003 Madrid, Spain

Francisco Gascón and Francisco J. Nieves[a)]
Departamento de Física Aplicada II, Universidad de Sevilla, E.T.S. de Arquitectura, Avenida Reina Mercedes, 2, 41012 Sevilla, Spain

(Received 5 November 2004; revised 4 March 2005; accepted 9 March 2005)

In this paper a method is proposed to characterize the elasticity of isotropic linear materials from the generation and detection of an acoustic surface wave. For the calculation of the elastic constants, it is sufficient that only one of the faces of the sample be accessible. The methodology is based on both the measurement of the Rayleigh wave velocity and on the determination of the normal to longitudinal amplitude ratio calculated from the normal and longitudinal components of the displacement of a point. The detection of two consecutive surface wave pulses using a single experimental setup permits the determination of the elastic constants. The method is applied to calculate Young's modulus and Poisson's ratio of an aluminum sample as well as their systematic uncertainties. The results obtained give a relative uncertainty for Young's modulus on the order of the sixth part of that calculated for Poisson's ratio. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1898663]

PACS numbers: 43.20.-f, 43.40.-r, 43.40.Yq, 43.40.At [YHB] Pages: 3469–3477

I. INTRODUCTION

Historically, two different techniques have been used to determine the elastic constants of a material, one static and the other dynamic. One nondestructive testing technique associated with dynamic methods is based on the measurement of bulk velocities; Young's modulus E and Poisson's ratio ν are then calculated from the longitudinal and shear ultrasonic wave velocities.[1] Measurement of resonance frequencies of a slender bar[2] also leads to the characterization of the elasticity of materials. The determination of the dynamic elastic constants from the measurement of the resonance frequencies of samples of diverse shape, such as cubes, parallelepipeds, or nonslender bars,[3–5] is a method which has been increasingly used. The resonance methods were limited by the lack of sufficiently exact solutions to the problem of estimating the frequencies of vibration of an elastic solid of arbitrary shape. In the aforementioned procedures, mainly piezoelectric transducers are used as generators and detectors of stress waves.

An alternative methodology, based upon the laser-ultrasonic technique, uses lasers to generate and to detect ultrasounds. One of the advantages of using this technique is that contact with the sample is not required, thereby avoiding the necessity for correction factors to be considered due to the added overweight. In addition, the detection is almost point-like and the detector has a wide bandwidth in its frequency response. Different contributions[6–9] describe the use of lasers as detectors and/or generators instead of piezoelectric transducers; the results provided show the extensive range of application of this technique.

Another method applied in order to evaluate the elastic constants is partly based on the measurement of the velocity of the surface acoustic wave or Rayleigh wave. In this setup the source and the detector are on the same face of the sample, which could be convenient in the case of complex geometric form or when only one single face is accessible. In addition, the identification of these waves is simple since they propagate great distance without hard attenuation. In effect, since propagation is two-dimensional rather than three-dimensional, the intensity is not inversely proportional to the square of the distance but to the distance itself. Rayleigh waves can be generated using a laser and detected with a point-receiver transducer[10] in order to measure the surface-wave energy velocities. The anisotropic elastic constants are then recovered by an optimization method such that the difference between the calculated and measured energy velocities reaches a minimum. In the work by Wu and Fang,[11] Rayleigh waves are generated by means of an impact and the resulting waves are detected with transducers. If the velocity of Rayleigh waves and that of P waves are known, the elastic constants of isotropic materials can be calculated. The generation of Rayleigh waves following the wedge method and detection by laser interferometry have been used to measure the velocity of Rayleigh waves in an anisotropic medium (slate).[12]

The quotient of velocity of Rayleigh waves and P waves is adequate for the calculation of Poisson's ratio. In fact, Poisson's ratio is a single-valued function of that quotient (see for instance Fig. 6.8 in Ref. 13). It is preferable to determine such velocities using two accessible faces. Despite the fact that Poisson's ratio has been calculated[11] from such velocities, using only one accessible face, if the distance traveled is short then the detected waves can be partially

a)Electronic mail: nieves@us.es

superposed, which causes deviations in the measurement of Poisson's ratio.

In this work, a direct method is proposed to determine the elastic constants of any isotropic material from the detection of Rayleigh wave. The method could be applied using only one single Rayleigh wave pulse, although for convenience more than one pulse is used here. The measurement of the material's density, the velocity of the surface wave, and the quotient of its normal and longitudinal displacement components are required. The key of this work is the sole use of Rayleigh waves. The detection at a point on the surface of a single surface wave pulse is enough to measure its velocity, as well as to calculate the normal to longitudinal displacement component ratio from the amplitudes of one harmonic component of the wave.

II. RAYLEIGH WAVES AND DYNAMIC ELASTIC CONSTANTS

Rayleigh waves[14] propagate along the free surface of a body without penetrating it. More specifically, the disturbance is confined to a surface layer whose depth is in the order of its wavelength. In homogeneous and isotropic media the particle motion is elliptical and retrograde near the surface, and the major normal axis of the ellipse is normal to the free surface, i.e., the vibration plane is perpendicular to the surface and contains the direction of propagation.[15] The velocity of Rayleigh wave v_R, and the phase velocities v_P and v_S of the bulk waves *P* and *S*, respectively, are related[16,17] by the equation

$$\eta^3-8\eta^2+8\eta\left(3-\frac{2v_S^2}{v_P^2}\right)-16\left(1-\frac{v_S^2}{v_P^2}\right)=0, \tag{1}$$

where

$$\eta=\left(\frac{v_R}{v_S}\right)^2. \tag{2}$$

In addition, the quotient of normal u_t and longitudinal u_l amplitude components of Rayleigh waves at a point on the surface, $a\equiv u_t/u_l$, is related to η by the formula

$$a\equiv\frac{u_t}{u_l}=\frac{2-\eta}{2\sqrt{1-\eta}}. \tag{3}$$

This quotient or its inverse, the ellipticity, is an important parameter which reflects fundamental properties of the elastic material and can be used in different applications in seismology and nondestructive testing.[18]

The experimental measurement of the normal and longitudinal amplitude components of the displacement of a point due to Rayleigh waves propagating along the surface of an isotropic material allows the calculation of its quotient *a*. Equation (3) gives η from *a*. With the values of η and v_R, Eq. (2) yields v_S. With the values of η and v_S, Eq. (1) leads to the calculation of v_P. Finally, v_S and v_P determine ν and *E*.

Since Poisson's ratio ν varies from 0 to 1/2 for ordinary materials, the quotient $v_S/v_P=\sqrt{(1-2\nu)/(2-2\nu)}$ ranges between $1/\sqrt{2}$ and 0 and the quotient η is between 0.7639 and 0.9126; therefore, *a* is in the range 1.2720 to 1.8394. Hence, $a>1$ and, furthermore, $\sqrt{a^2-1}<a$. From Eq. (3) the only positive solution obtained is

$$\eta=-2(a^2-1)+2a\sqrt{a^2-1}$$
$$=-2\left[\left(\frac{u_t}{u_l}\right)^2-1\right]+2\frac{u_t}{u_l}\sqrt{\left(\frac{u_t}{u_l}\right)^2-1}. \tag{4}$$

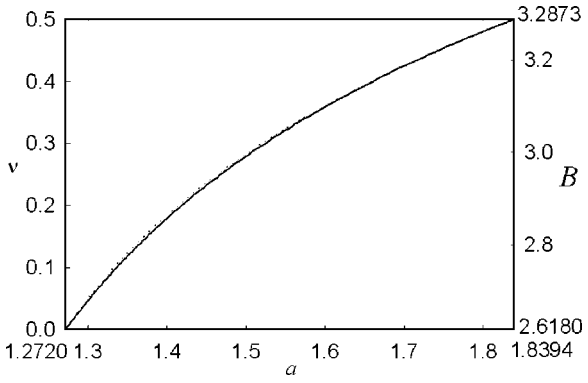


FIG. 1. Poisson's ratio ν (continuous line) and nondimensional quantity *B* (dotted line) versus quotient of amplitude $a\equiv u_t/u_l$.

Therefore, once η is calculated from *a* by means of Eq. (4), the experimental measurement of v_R determines the velocity $v_S=v_R/\sqrt{\eta}$ using Eq. (2). Since Eq. (4) provides a unique solution for η, Eq. (1) gives a single value for v_P

$$v_P=4v_R\left[\frac{1-\eta}{\eta(16-24\eta+8\eta^2-\eta^3)}\right]^{1/2}. \tag{5}$$

The well-known relation between Poisson's ratio ν and the velocities v_P and v_S, together with Eqs. (2) and (5), leads to the expression

$$\nu=\frac{\frac{1}{2}\left(\frac{v_P}{v_S}\right)^2-1}{\left(\frac{v_P}{v_S}\right)^2-1}=\frac{\eta^3-8\eta^2+16\eta-8}{\eta^3-8\eta^2+8\eta}. \tag{6}$$

Equations (6) and (4) yield ν. Note that Poisson's ratio only depends on η; therefore, the quotient u_t/u_l determines ν.

Equation (2) and the well-known relation of v_S with the elastic constants allow Young's modulus *E* to be written as

$$E=2(1+\nu)\rho v_S^2=\frac{4\rho v_R^2}{\eta^2}\frac{\eta^3-8\eta^2+12\eta-4}{\eta^2-8\eta+8}. \tag{7}$$

Equation (7) and Eq. (4), together with the values of v_R and density ρ, are sufficient to calculate *E*.

Therefore, it is demonstrated that the measurements of u_t, u_l, v_R, and ρ allow the calculation of the two dynamic elastic constants, *E* and ν.

Figure 1 graphically represents Eq. (6) (continuous line), by taking Eq. (4) into account. That is to say, Poisson's ratio ν for the ordinary materials is given as a function of the quotient of amplitudes *a*. In the same figure the nondimensional quantity

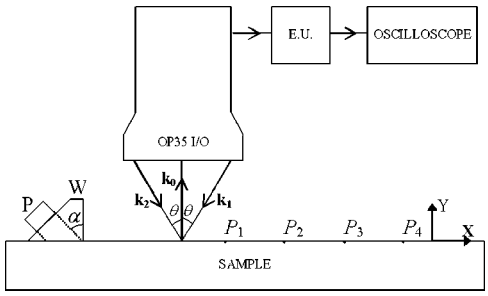


FIG. 2. Experimental setup. Rayleigh wave is generated using the wedge *W* and it propagates along the surface of the sample. The detection of the components of the displacement is made with an optical interferometer, whose output signal is demodulated in the electronic unit (E.U.), and finally, it is visualized by means of an oscilloscope.

$$B \equiv E/\rho v_R^2, \tag{8}$$

obtained from Eqs. (7) and (4), is shown (dotted line). Although both functions seem to be totally independent, it could be concluded from the analysis of the figure that a relationship seems to exist between them. The approximate relation deduced is

$$B = 2.618 + 1.332\nu. \tag{9}$$

An analytical explanation has not been found for such a peculiar relation. However, a graphical application is deduced. In effect, once the quotient of amplitudes a is calculated, a vertical straight line drawn through the value of this quotient intersects the graph at a point. A horizontal line through this point provides the value of Poisson's ratio to the left and the nondimensional number B to the right. Young's modulus is calculated from Eq. (8).

The aspect ratio a of the elliptical trajectory of the surface particle is of great importance. Such a ratio is used here to calculate the elastic constants and it is equal to both the ratio between the vector and scalar potential amplitudes and the ratio between the shear and normal stresses in any plane parallel to the surface at any depth.[19]

III. EXPERIMENTAL PROCEDURE

Our methodology, based upon the determination of the Rayleigh wave velocity v_R and the quotient of amplitude a, is applied in order to calculate the elastic constants of a commercial aluminum block, with density $\rho = 2827$ kg/m^3. This block is previously annealed in order to eliminate its internal stresses; the sample tested is then homogeneous. Figure 2 presents the experimental setup. The excitation of Rayleigh waves is performed by means of the wedge method[15] based on the conversion of *P* waves into Rayleigh waves. A piezoelectric transducer of 0.5 MHz is set on the sloping surface of the wedge, which is acoustically coupled through one of its faces to the surface of the sample. The *P* wave emitted by the transducer propagates through the wedge and encounters the interface between the wedge and the sample at the incident angle α. The generation of the surface wave is optimal when the angle of incidence is equal to the optimum angle α_c, where $\alpha_c = \sin^{-1}(v_{Pw}/v_R)$, and v_{Pw} is the *P*-wave velocity in the wedge. v_{Pw} must be less than v_R in order for this technique to work. A methacrylate wedge is adhered to the sample. This material is chosen because the velocity of the *P* waves in methacrylate is 2730 m/s, lower than the *P*-wave velocity in aluminum, and hence the condition for total reflection is satisfied. The angle of the wedge must be close to the optimum angle in order to obtain optimum generation; an angle of 70 deg is used as a trial in our experiments.

Rayleigh waves are detected by means of an optical heterodyne interferometer OP35-I/O from Ultra-Optec,[20] which allows the point-like detection of the two components of the wave at the same point, although sequentially. The detection method is based on the phenomenon of speckle associated with the random distribution of maxima and minima of intensity observed when coherent light is scattered by a scratched surface. This detection system is equipped with a 5-mW He–Ne laser and has a bandwidth between 1 kHz and 35 MHz and a sensitivity of approximately 1 nm. The laser beam crosses an acoustic-optical modulator (Bragg cell) which yields two outcoming beams, one being undeflected and unshifted in frequency and the second one being deflected and frequency shifted by 40 MHz.

It is expected that the wavelength of the Rayleigh wave is large compared to the diameter (20 μm) of the zone illuminated by the laser on the surface of the sample. Therefore, detection can be regarded as "point-like."

When trying to determine the tangential component (in plane or longitudinal), both beams are directed so that they strike the same point of the sample surface, where the directions of incidence $\mathbf{k_1}$ and $\mathbf{k_2}$ are symmetrical with respect to the normal to the surface, as shown in Fig. 2. The angle between $\mathbf{k_1}$ and the normal and the angle between $\mathbf{k_2}$ and the normal are both approximately equal to $\theta = 37.5$ deg in our case. The scattered light is collected in the normal direction $\mathbf{k_0}$ and directed to the detector.

If the objective is the detection of the normal component (out of plane), then only the frequency-shifted beam is incident on the surface in direction $\mathbf{k_1}$ and the deflected light is collected along $\mathbf{k_2}$. It is then directed to the beam mixer, where it interferes with the unshifted frequency beam; finally, the resulting signal arrives at the detector.

A Rayleigh wave propagates on the surface of the sample, as a result of the conversion of the *P* wave. Let axis *OX* be in the direction of propagation and axis *OY* perpendicular to the surface. Therefore, the total displacement $\mathbf{u}$ of a point can be expressed by its component on the surface of the sample u_l (in plane) and the perpendicular component u_t (out of plane), i.e., $\mathbf{u} = u_l\mathbf{i} + u_t\mathbf{j}$. This assembly permits the detection of the longitudinal u_l and transverse u_t components of the motion of the point by sequentially using the in-plane and out-of-plane detection configurations. This is done by varying the paths of the beams within the interferometer, an elementary operation carried out by means of a switch which opens and closes shutters. If the system is operating in the in-plane mode and $\mathbf{u}$ is the displacement of a point on the surface, the resulting variation of the phase associated with the path difference is equal to $\Delta_l = [4\pi u_l(t)\sin\theta]/\lambda$, where λ is the wavelength of the laser and t the time. In the same way, for the out-of-plane configuration, such a variation is equal to $\Delta_t = [4\pi u_t(t)\cos\theta]/\lambda$.

The ac signal obtained as output of the optical unit is

demodulated in the electronic unit to obtain a potential difference proportional to the displacement of the point of detection on the surface. The factor of calibration of the system for the range of frequencies considered is approximately 49.2 and 64.1 nm/V for the transverse and longitudinal components, respectively. The quotient of the potential difference V_t', associated with the transverse component, and that denoted V_l', corresponding to the longitudinal component, is $V_t'/V_l' = (u_t \cos\theta)/(u_l \sin\theta)$, which gives $u_t/u_l = (V_t'/V_l')\tan\theta$. Hence, it is not necessary to consider the calibration factor. The generator which supplies the piezoelectric transducer simultaneously sends an ultrasonic pulse and triggers the digital oscilloscope which measures V_t' as a function of time. The generator then triggers another equal pulse and V_l' is synchronously measured with the previous pulse. The effect is as if V_t' and V_l' were measured simultaneously, except for the noise. The quotient of the transverse and longitudinal components can be obtained from both the quotient of the potential differences and the tangent of the angle created by the incident beam and the perpendicular to the surface, i.e., $u_t/u_l = (V_t'/V_l')\tan\theta$. Indeed, in order to measure the potential differences accurately, it is more advisable to generate not a unique pulse but a series of repeated identical pulses and to average the detected values.

Using this optical interferometer the Rayleigh wave velocity is also determined. In theory, the velocity v_R is calculated by measuring the separation of the emitter from the point of detection. This distance divided by the time elapsed is the Rayleigh wave velocity. However, the emitter is not point-like; hence, the separation of the point of detection to the emitter is not well defined. In addition, there is always a certain delay between the time of application of the tension pulse to the transducer and that when the resulting wave leaves its external face in contact with the wedge. These facts lead us to look for an alternative method. A simple and precise method is based upon the determination of the elapsed time, t_i, of the wave at different points located at coordinates x_i. In Fig. 2, P_1, P_2, P_3, and P_4 represent the points of detection where the arrival of the waves is recorded. The slope of the straight line of linear regression of the set of values (x_i, t_i) yields the velocity v_R.

IV. EXPERIMENTAL RESULTS

For the observation and analysis of the signal detected in the oscilloscope, the sampling frequency used is 100 Msa/s and the number of points $N = 2500$. The signals are averaged in order to improve the signal-to-noise ratio. The peak-to-peak voltage for the normal component measured by the oscilloscope is $V_t' = 62.8$ mV and for the tangential component, $V_l' = 32.8$ mV. In order to turn the volts measured by the oscilloscope into meters, the detected potential differences are multiplied by the corresponding factor of calibration. Figure 3 shows the results obtained for the displacement associated with the normal component (measured in nm and drawn as a continuous line) and longitudinal component (in nm, dotted line) of the Rayleigh wave at point P_l, previously shown in Fig. 2, as approximately 7 cm from the end of the wedge. In Fig. 3, the normal component has oscillations of high frequency attributable to the noise of magnitude similar to the magnitude of the signal until $t \approx 8\,\mu$s. For the longitudinal component, the noise is the only signal until $t \approx 10\,\mu$s; thus, there is no useful information before those values are reached. It can be verified from $t \approx 10\,\mu$s that when a harmonic plane Rayleigh wave propagates, the transverse component is shifted 90 deg in phase with respect to the longitudinal component. In Fig. 4 the resulting trajectory described by the analyzed point is presented.

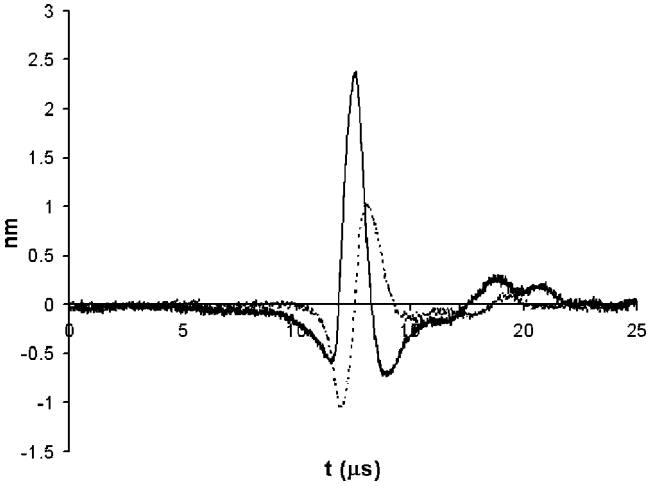


FIG. 3. Normal component (solid line) and tangential component (dotted line) of the displacement in nanometers of the Rayleigh wave detected at P_1.

The detected signals V_t' and V_l' are later smoothed with a Hanning window, which has a Gaussian distribution. The peak of the out-of-plane component coincides with the center of the time window. Therefore, the signals have consistent amplitude at the center and zero amplitude at each end. The surface wave components essentially retain the same amplitude, whereas noise components are reduced. Figure 5 shows the fast Fourier transform of the measured voltages which were smoothed with a Hanning window. The harmonic components shown in the interval of frequencies of apparent in-

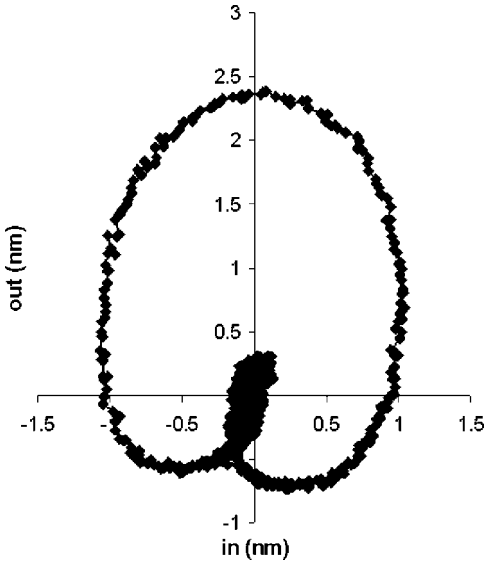


FIG. 4. Trajectory described by the point P_1 due to the pulse.

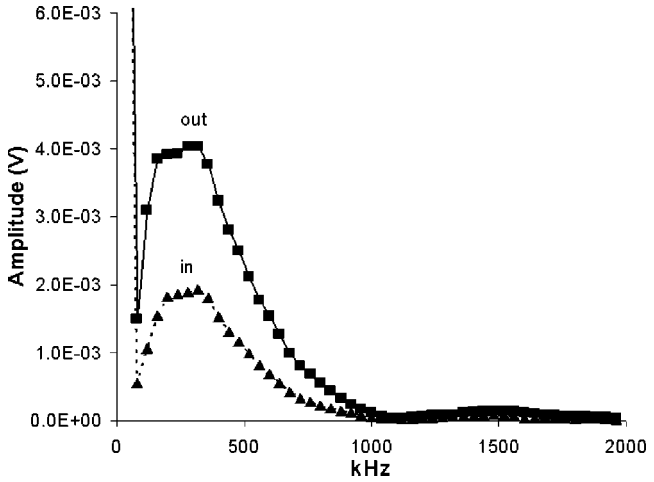


FIG. 5. Amplitudes obtained from the calculation of the FFT of the potential differences detected for the transverse and longitudinal components. The interval of frequencies of interest around the central frequency of the transducer is shown.

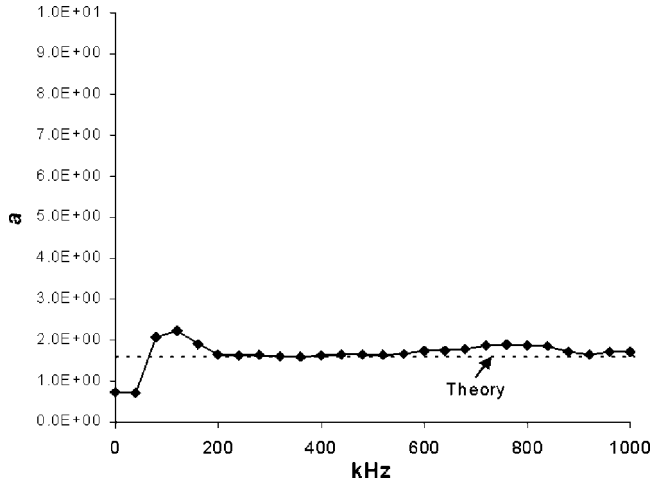


FIG. 6. Ratio between the normal and tangential amplitude components of the displacement of the Rayleigh wave for the interval of frequencies of interest.

terest (i.e., frequencies close to 0.5 MHz, which is the central frequency of the piezoelectric transducer used as wave generator) are transverse and longitudinal. Hence, it is expected that the harmonic components of a frequency of approximately 0.5 MHz would be of maximum amplitude, although the maximum amplitude does not correspond to 0.5 MHz, but to 320 kHz. The amplitudes given by the FFT for the latter frequency are 4.04 and 1.94 mV for the transverse and longitudinal components, respectively. By numerically studying the propagation of a damped harmonic wave, it is deduced that the maximum amplitude in the spectrum corresponds to a frequency smaller than that theoretically obtained without damping. The fact that the wave undergoes damping when being propagated through the methacrylate wedge and, to a lesser extent, through the aluminum sample, could contribute towards this shift in the maximum amplitude. It could also contribute towards such a shift the conversion of the *P* wave into a Rayleigh wave since, for smaller frequencies, this conversion seems more efficient. As shown in Ref. 21, to effectively excite the Rayleigh surface wave, the source frequency should be controlled in the proper range.

With the aim of detecting, by means of the interferometer, the direct wave emitted by the transducer, it is adhered to a face of the aluminum plate, whose thickness is 5 cm, and the displacement of a point is detected on the parallel face, the detection point being opposite the central point of the transducer. The spectrum of this signal shows that the interval of frequencies, for an amplitude half of the maximum, corresponds to a bandwidth between 150 and 725 kHz. This result seems to confirm that the maximum of amplitude shifts to a lower frequency. In addition, a second experiment with the transducer is performed. A scratched aluminum foil is adhered on the exciting face of the transducer in order to measure with the laser interferometer the displacements at a point on its face. The maximum amplitude frequency associated with the measured out-of-plane displacement is 427 kHz, the bandwidth being 178 kHz (from 372 to 550 kHz), which shows that the down shift of the frequency is also due to the transducer itself.

Figure 6 represents the quotient of amplitudes versus the frequency for the amplitudes shown in Fig. 5. The quotient a, shown in Fig. 6, is obtained, for each frequency, by determining the quotient of the amplitudes given by the FFT and multiplying the result by $\tan\theta$. Moreover, Fig. 7 shows the difference of phase between the transverse and longitudinal components versus the frequency. From the analysis of these figures it is deduced that both the quotient a and the difference of phase remain approximately constant in the interval of frequencies between 200 and 900 kHz. It seems convenient for the calculation of the elastic constants to take the quotient corresponding to the frequency whose amplitude in the spectrum is at a maximum, i.e., the quotient for 320 kHz, whose value is $a=1.60$. The difference of phase obtained between the transverse and longitudinal components for that frequency is 89.3°. The displacements associated with the amplitudes given by the FFT are calculated by multiplying the potential differences for 320 kHz by the factor of calibration corresponding to each component; the values 0.199 and 0.124 nm are obtained for the transverse and longitudi-

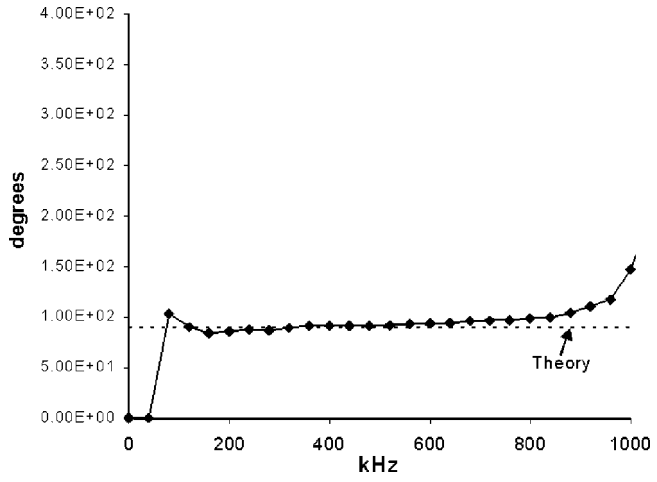


FIG. 7. Phase difference between the normal and longitudinal components.

nal components, respectively. Similar graphs to Fig. 6 and Fig. 7, for the quotient of amplitudes and difference of phase as a function of the frequency, are obtained for all the points of detection.

The Rayleigh wave velocity is then determined from the arrival time of the wave to four positions. The sample is mounted on a translational mechanism with which it can be moved with a precision of 0.01 mm. On fitting a straight line to the measured data (x_i, t_i), the straight line obtained is $y = 2868.2x - 9 \cdot 10^{-6}$, $r = 1.000$, whose slope gives the wave velocity $v_R = 2868$ m/s.

Once both $a = u_t/u_l$ and v_R are known, it is possible to determine Young's modulus and Poisson's ratio as analyzed in Sec. II. By taking the value of the ratio a corresponding to the frequency whose amplitude is at a maximum, i.e., $a = 1.60$, Eq. (4) gives $\eta = (v_R/v_s)^2 = 0.877$. From the value η, Eq. (6) directly provides Poisson's ratio, $\upsilon = 0.359$. In addition, since v_R and ρ are known, Eq. (7) gives Young's modulus, $E = 72.1$ GPa.

In order to verify the validity of these results, the bulk wave velocity v_P is determined by using two *P*-wave ultrasonic transducers adhered to opposite faces of the sample, one acting as emitter and the other as receiver. The velocity obtained is $v_P = 6345$ m/s. Repeating the measurement with *S*-wave transducers, $v_S = 3092$ m/s is obtained. With these values, the Rayleigh wave velocity calculated by applying Eq. (1) is $v_R = 2888$ m/s. Taking the values v_P and v_S as input data, E and ν are 72.7 GPa and 0.344, respectively. The relative differences of the values obtained from the Rayleigh wave and the previous values from the bulk velocities are 0.82% for E and 4.4% for ν.

These differences between the values of the elastic constants can be due to diverse factors affecting the processes of generation and detection. First, small deviations in the measurement of a imply important variations of the value of ν, as deduced from Fig. 1. Another factor is the low efficiency of generation: in the experiences carried out, the value of the amplitude of the Rayleigh wave approaches the limit of detection of the system, and therefore noise can mask measurement. This is especially true when considering the in-plane component. What is more, the angle of the wedge decisively influences the amplitude, since such an angle must be the optimum; in our case $\theta_c = \sin^{-1}(2730/2868) = 72.15$ deg, which is very close to the 70 deg of the wedge used. In addition, the polarization plane of the wave must coincide with the plane of detection, otherwise the in-plane component would be only partially detected.

The results obtained for the Rayleigh wave show that the field of this wave is not a perfect plane wavefront. The peak-to-peak voltages measured at points P_1, P_2, P_3, and P_4 differ up to 2.9% for the out-of-plane component and up to 12% for the in-plane component. Fluctuations in the ratio u_t/u_l are observed at the different points of detection, whereby the variations in the values obtained approach 10%. Close values have been obtained by other researchers with similar systems of detection.[20] The causes which could account for these discrepancies include instability of the detection system; changes in the connection with the transducer or in the pulse emitted during the averaged time; the effect of the damping applied to the piezoelectric device; and the absorption of the wave by the material.

As regards the wavelength of the Rayleigh wave, $\lambda = v_R/f$, in the present study its value is equal to approximately 5 mm, which is much larger than the diameter of the illuminated area (0.02 mm). Hence, the detection can be considered to be point-like.

V. ACCURACY

The standard uncertainty, limited to the component of uncertainty arising from a systematic effect, is evaluated for each measurement and for each involved quantity. The values obtained represent an estimation of the corresponding uncertainties, and they are not based on any statistical method.[22] Poisson's ratio ν and Young's modulus E are calculated from indirect measurements applying Eqs. (6) and (7), respectively. It is generally accepted that the systematic uncertainty of an indirect measurement is estimated by means of the differential of the function which relates the magnitude to be calculated to the directly measured magnitudes.[23]

The potential differences associated with the transverse and longitudinal components are directly measured in the experiments carried out. The digital oscilloscope used, Tektronic TDS-430A, provides 9 bits for a sampling frequency of 100 Msa/s and 2500 points. The detection of signals is performed with the vertical scale of the oscilloscope corresponding to 80 mV. With these values, a sensitivity of $U_V = 0.156$ mV is obtained for the potential differences recorded on the oscilloscope.

In the calculation of ν, the values of the amplitudes given by the FFT of the signals registered by means of the oscilloscope are used. Therefore, it is necessary to determine the uncertainty which corresponds to these amplitudes.

The problem to solve now is the calculation of the uncertainty of the Fourier transform of a sampled function, i.e., of a DFT. The digital oscilloscope used in the experiments and described in this article takes N samples whose values are f_j, $\forall j = 0,1,2,...,N-1$. This sampled function can always be written as a sum of complex harmonics,[24] for example, in the form

$$f_j = \sum_{k=0}^{N-1} F_k \exp(i2\pi kj/N), \tag{10}$$

where i is the imaginary unit and F_k, $\forall k = 0,1,2,...,N-1$, is the so-called discrete Fourier transform of the sampled function f_j, whose expression is given by

$$F_k = N^{-1} \sum_{j=0}^{N-1} f_j \exp(-i2\pi kj/N). \tag{11}$$

In the present case f_j is real and its "true" value is included in the interval $f_j \pm U_{f_j}$. If the most unfavorable case is chosen, Eq. (11) permits the estimation of the maximum value of the kth harmonic

$$F_k + U_{F_k} = \max\{N^{-1}[(f_0 \pm U_{f_0})\exp(0) + (f_1 \pm U_{f_1}) \times \exp(-i2\pi k/N) + \cdots]\}$$
$$= N^{-1}\{f_0 \exp(0) + f_1 \exp(-i2\pi k/N) + \cdots + \max[\pm U_{f_0}\exp(0) \pm U_{f_1} \times \exp(-i2\pi k/N) \pm \cdots]\}. \quad (12)$$

Hence

$$U_{F_k} = \max\{N^{-1}[\pm U_{f_0}\exp(0) \pm U_{f_1} \times \exp(-i2\pi k/N) + \cdots]\}$$
$$= N^{-1}\max\left\{\sum_{j=0}^{N-1}[\pm U_{f_j}\exp(-i2\pi kj/N)]\right\}. \quad (13)$$

As all measurements are made in the same oscilloscope range, U_{f_j} is unchanged for all measurements, and by denoting U_{f_j} as U_f, it follows that

$$U_{F_k} = N^{-1}U_f \max\left\{\sum_{j=0}^{N-1}[\pm\exp(-i2\pi kj/N)]\right\}. \quad (14)$$

The terms in the sum are homogeneously distributed over the complex plane. Let the first term of this sum have a positive sign. Therefore, the sum of the complex values is at a maximum if the sign assigned to each term causes the associated complex number to be a positive real part. When representing the complex numbers taking into account the suitable sign to maximize the sum, it is found that the maximum value of the modulus of the sum is

$$U_{F_k} = N^{-1}U_f \sum_{j=0}^{N-1}|\cos(2\pi kj/N)|. \quad (15)$$

The value of the sum for large values of N multiplied by N^{-1} is equal to $2/\pi$, which is the result for almost all values of k. This fact leads to

$$U_{F_k} = \frac{2}{\pi}U_f. \quad (16)$$

In our case $U_f = 0.156\,\text{mV}$, which is the value U_V previously obtained for the sensitivity of the potential differences of both operation modes of the interferometer in the process of digitalization. Finally, in the range of frequencies of interest, a systematic uncertainty $U_F = 0.0995\,\text{mV}$ is deduced.

In order to estimate the uncertainty from the angle, the data from the manufacturer are used and the two incidence angles are supposed equal. The angles θ are both equal to 37.5 deg and their estimated uncertainty, 0.026 radians.[20]

The next objective is to evaluate the uncertainty of $a = u_t/u_l = (V_t'/V_l')\tan\theta$, which is equal to

$$U_a = \left(\frac{1}{V_l'} \times \tan\theta\right)(U_F) + \left(\frac{V_t'}{V_l'^2} \times \tan\theta\right)(U_F) + \left(\frac{V_t'}{V_l'}\frac{1}{\cos^2\theta}\right)(U_\theta). \quad (17)$$

As mentioned earlier, for the case under study the maximum values of the amplitudes in the spectrum for the frequency 320 kHz are equal to $V_t' = 4.04$ and $V_l' = 1.94\,\text{mV}$. For these values, $a = 1.60$ and its uncertainty $U_a = 0.208$. The first term of Eq. (17), associated with the transverse component, represents approximately 19% of the value of U_a, whereas the contribution of the second term corresponding to the longitudinal component is 39.5%, and that of the term due to the measurement of the angle, 41.5%.

Analogously, the absolute value of the differential of η from Eq. (4) is calculated in order to determine U_η, i.e., $U_\eta = |\partial\eta/\partial a|U_a$, and the values obtained are $\eta = 0.410$ and $U_\eta = 0.041$. Finally, the same methodology applied to Eq. (6) gives $U_\nu = |\partial\nu/\partial\eta|U_\eta$, whose value is $U_\nu = 0.150$, and relative value $U_\nu/\nu = 41.7\%$. If it were assumed that the angle could be measured exactly, the relative value U_ν/ν, due to electric quantities, would be approximately 24%.

The systematic uncertainty of Young's modulus E is estimated with the help of Eq. (7). In order to measure the mass $m = 5.077\,\text{kg}$ of the sample, a scale is used with a sensitivity $U_m = 1\,\text{g}$ and the volume is determined from the dimensions of the parallelepiped, $L_1 = 30.0\,\text{cm}$, $L_2 = 119.71\,\text{mm}$, $L_3 = 50.00\,\text{mm}$, whose respective sensitivities are $U_{L_1} = 1\,\text{mm}$, $U_{L_2} = U_{L_3} = 0.05\,\text{mm}$. The density obtained is $\rho = 2827\,\text{kg/m}^3$ and its uncertainty $U_\rho = 13.99\,\text{kg/m}^3$. The Rayleigh wave velocity turns out to have an uncertainty of $U_{vR} = 6.68\,\text{m/s}$ since, in the measurement of the time elapsed, it is estimated that $U_T = 0.01\,\mu\text{s}$ and, for the measurement of the distance between the points of detection, the translational mechanism is used with $U_d = 0.01\,\text{mm}$. Finally, a systematic uncertainty for E of value $U_E = 5.26\,\text{GPa}$ is obtained, which represents a relative value $U_E/E = 7.3\%$. This value is approximately the sixth part of that calculated for ν.

The estimation carried out for the uncertainty of the measurement of Poisson's ratio is valid for the material under examination and the experimental setup used. This estimation can be extended to most metals whose values of ν are similar. If the uncertainties of the elastic constants of other materials were calculated, the results would be different, even with the same experimental setup. The uncertainty of ν and E can also be estimated by considering average values in the calculations.[25] In effect, the quotient $a = u_t/u_l$ varies nonlinearly with ν; hence by taking ν as a function of a, which is considered as an independent variable, the differential of ν is $d\nu = (d\nu/da)da$. For a specific uncertainty U_a, the uncertainty of ν will be maximum when the derivative of ν with respect to a is maximum. However, as $U_\nu \approx (\Delta\nu/\Delta a)U_a$ and the interval of existence of ν is $\Delta\nu = 0.5$ and that of a, $\Delta a = 0.56$, an average estimation can be made for all the materials, where $U_\nu = 0.881\,U_a$. If this result is applied to the case studied with the aluminum sample, it gives $U_\nu = 0.183$, which is a result whose order of magnitude differs slightly from that previously calculated. In order to estimate the uncertainty of E, the uncertainty of $B = E/\rho v_R^2$ is first evaluated. If it is assumed that ρ and v_R are measured with high accuracy, then $U_E = \rho v_R^2 U_B$. In the same way, $U_B \approx (\Delta B/\Delta a)U_a$. Hence, by taking the values of ΔB and Δa given by Fig. 1, it is concluded that $U_B = 1.18U_a$, and hence $U_E = 1.18\rho v_R^2 U_a$. For the case studied and the experi-

mental setup used, $U_E = 5.70$ GPa is obtained, which is a result close to that previously calculated.

The presented methodology could be called absolute. Nevertheless, a relative measurement or calibration measurement could also be done. There are at least two similar ways of doing such a calibration using a test sample whose elastic constants have been measured by a very accurate method (but with error), for instance by pulse–echo. (1) Equation (6) determines η from ν. Thereafter, Eq. (4) permits calculation of u_t/u_l, by substitution of η. If V_t'/V_l' is measured for a Rayleigh wave on the surface of the reference sample, then the angle θ can be calculated by the formula $\tan\theta = (u_t/u_l)/(V_t'/V_l')$. However, the value of θ will be affected by their corresponding errors, which should be substituted into Eq. (17). The results obtained from the Rayleigh wave along with those from the pulse–echo method are given in Sec. IV. From such results for the sample studied and following the above methodology, it is found that the value of θ is 37.1 deg. Hence, the elastic constants for the tested sample from the Rayleigh wave measurements and this value of θ are $E = 71.53$ GPa and $\nu = 0.341$. (2) One could maintain θ as an unknown value and measure V_t' and V_l' for two test samples, one with unknown elastic moduli and another with known elastic constants for reference purposes. For both samples, it is calculated that $u_t/u_l = (V_t'/V_l')\tan\theta$ and the respective expressions of η in function of θ, by Eq. (4). Finally, the respective substitutions into Eq. (6) give two equations, which permit the calculation of the unknown ν. Both procedures are affected by the errors of four voltage measurements, and consequently, they do not seem to offer much better than the method that includes the goniometric measurement of θ and two voltages.

VI. CONCLUSIONS

In this work the possibility of characterizing the elasticity of isotropic materials from measurements taken on its surface is demonstrated. This methodology could be applied in the study of materials where only one single face is accessible. Poisson's ratio is calculated from the quotient of the transversal and longitudinal components of the Rayleigh surface wave, whereas Young's modulus is obtained from this ratio together with the Rayleigh wave velocity. The method is based upon the use of speckle interferometry, which permits the detection of the two components of the displacement of the surface wave at a given point. The detection is point-like and without contact, thereby avoiding the necessity of introducing correction factors.

With such a method, the main difficulties are related to the generation of the Rayleigh wave. This methodology could be improved by designing a wedge whose slope could be varied in a continuous form in order to match the incident angle to the optimum angle, thereby increasing the signal-to-noise ratio and therefore carrying out the experiment in optimal conditions. In addition, other methods of generation of Rayleigh waves which do not require critical incidence could be applied, such as the laser ultrasonic technique and generation by means of an impact. That is to say, the same methodology could give more precise results by generating a wave of larger amplitude compared with the noise. Hence, the high uncertainty obtained for Poisson's ratio is due, among other factors, to the low efficiency of generation.

Furthermore, the mathematical dependence of ν on the amplitude quotient is such that its accuracy is critical in relation to the measurements carried out, which makes the method difficult to apply. Thus, a relative uncertainty of such a quotient of 13% yields a value for the uncertainty of ν on the order of 42%. In order to have a low uncertainty of ν, it is necessary to measure the amplitude of the wave with high accuracy. On the other hand, the method provides much more precise values of E. It is estimated, taking average values, that the uncertainty of ν is approximately six times that of E,[25] which demonstrates the existing difficulty of measuring ν with high accuracy when applying this method.

A classic method for the calculation of ν is the measurement of v_P and v_S through a material plate. From the equation $v_p/v_s = \sqrt{(1-2\nu)/[2(1-\nu)]}$, it is deduced that $U_\nu = |\partial\nu/\partial(v_P/v_S)|U(v_P/v_S)$, and as the interval of existence of v_p/v_s is (0,0.71), the derivative $|\partial\nu/\partial(v_P/v_S)|$ is approximately $0.5/0.71 = 0.7$. Since the values obtained for the velocities are $v_P = 6345$ m/s and $v_S = 3092$ m/s, which are affected by their respective uncertainties $U_{v_P} = 4.63$ m/s and $U_{v_S} = 1.26$ m/s, the classic method yields a systematic uncertainty for Poisson's ratio equal to $U_v = 0.0016$ and a relative value $U_\nu/\nu = 0.5\%$. Despite being much smaller than that corresponding to the proposed method, this method has the serious disadvantage that the test sample should have preferably two opposite faces accessible, whereas the method proposed does not require more than one single accessible face.

ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Governmental Plan of Scientific Research, Development and Technological Innovation (*Plan Nacional de I+D+i 2004–2007*) under Project No. BIA2004-07428-C02 of the Program *P.N. de Construcción*.

[1] E. P. Papadakis, *Physical Acoustics* (Academic, New York, 1976), Vol. XII, Chap. 5, pp. 277–374.
[2] C848-88 (1999), "Standard Test Method for Young's Modulus, Shear Modulus, and Poisson's Ratio For Ceramic Whitewares by Resonance," C623-92(2000) "Standard Test Method for Young's Modulus, Shear Modulus, and Poisson's Ratio for Glass and Glass-Ceramics by Resonance" and other ASTM standards (American Society for Testing and Materials, Philadelphia, PA).
[3] O. L. Anderson, "Rectangular parallelepiped resonance. A technique of resonance ultrasound and its applications to the determination of elasticity at high temperatures," J. Acoust. Soc. Am. **91**, 2245–2253 (1992).
[4] H. H. Demarest, "Cube-resonance method to determine the elastic constants of solids," J. Acoust. Soc. Am. **49**, 768–775 (1971).
[5] M. Senoo, T. Nishimura, A. Matsumuro, and S. Ishigau, "Measurement of the elastic constants of polycrystals in a cylindrical specimen by the resonance method," JSME Int. J. **30**, 747–753 (1987).
[6] J. D. Aussel and J. P. Monchalin, "Precision laser-ultrasonic velocity measurement and elastic constant determination," Ultrasonics **27**, 165–177 (1989).
[7] B. Castagnede, K. K. Kim, and W. Sachse, "Determination of the elastic constants of anisotropic materials using laser-generated ultrasonic signals," J. Appl. Phys. **70**, 150–157 (1991).
[8] K. Y. Kim, W. Sachse, and A. G. Every, "On the determination of sound speeds in cubic crystals and isotropic media using a broadband ultrasonic

point-source/point-receiver method,'' J. Acoust. Soc. Am. **93**, 1393–1405 (1993).
[9] A. Bayón, A. Varadé, and F. Gascón, ''Determination of the elastic constants of isotropic solids by optical heterodyne interferometry,'' J. Acoust. Soc. Am. **96**, 2589–2592 (1994).
[10] J. F. Chai and T. T. Wu, ''Determination of anisotropic elastic constants using laser-generated surface waves,'' J. Acoust. Soc. Am. **95**, 3232–3241 (1994).
[11] T. T. Wu and J. S. Fang, ''A new method for measuring in situ concrete elastic constants using horizontally polarized conical transducers,'' J. Acoust. Soc. Am. **101**, 330–336 (1997).
[12] A. Varadé, A. Bayón, P. N. J. Rasolofosaon, and F. Gascón, ''Experimental results on bulk waves and Rayleigh waves in slate,'' J. Acoust. Soc. Am. **99**, 292–298 (1996).
[13] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1975), p. 326.
[14] J. W. S. Rayleigh, ''On waves propagating along the plane surface of an elastic solid,'' Proc. London Math. Soc. **17**, 4–11 (1887).
[15] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[16] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, Oxford, 1970), pp. 111–112.
[17] D. Nkemzi, ''A new formula for the velocity of Rayleigh waves,'' Wave Motion **26**, 199–205 (1997).
[18] P. G. Malischewsky and F. Scherbaum, ''Love's formula and H/V-ratio (ellipticity) of Rayleigh waves,'' Wave Motion **40**(1), 57–67 (2004).
[19] W. Hassan and P. B. Nagy, ''Simplified expressions for the displacements and stresses produced by the Rayleigh wave,'' J. Acoust. Soc. Am. **104**, 3107–3110 (1998).
[20] J. P. Monchalin, J. D. Aussel, R. Heon, C. K. Jen, A. Boundreault, and R. Bernier, ''Measurements of in-plane and out-plane ultrasonic displacements by optical heterodyne interferometry,'' J. Nondestruct. Eval. **8**, 121–132 (1989).
[21] B. Zhang, M. Yu, C. Q. Lan, and W. Xiong, ''Elastic wave and excitation mechanim of surface waves in multilayered media,'' J. Acoust. Soc. Am. **100**, 3527–3538 (1996).
[22] B. N. Taylor and C. E. Kuyatt, ''Guidelines for Evaluating and Expressing the Uncertainties of NIST Measurement Results,'' NIST Technical Note 1297, U.S. Government Printing Office, Washington, D.C. (1994).
[23] J. R. Taylor, *An Introduction to Error Analysis* (University Science, Mill Valley, 1982), pp. 70–74.
[24] R. N. Bracewell, *The Fourier Transform and its Applications* (MacGraw-Hill, Tokyo, 1978).
[25] F. Gascón, A. Bayón, and F. J. Nieves, ''On the calculation of the elastic constants in isotropic materials from Rayleigh wave,'' Eleventh International Congress on Sound and Vibration, St. Petersburg, Russia (2004).

Phase-space analysis of acoustics fields and its application to waveguide

B. Teyssandier, V. Pagneux, L. Simon, and C. Depollier
Laboratoire d'Acoustique de l'Université du Maine, UMR-CNRS 6613, Université du Maine, Avenue Olivier Messiaen 72085 Le Mans Cedex 09, France

(Received 27 October 2003; revised 17 February 2005; accepted 6 March 2005)

A description of two-dimensional acoustic fields by means of a joint "space-wave number" representation is discussed. A function defined in the phase-space domain (x,y,k_x,k_y) is associated with a signal which is a function of spatial coordinates (x,y). This paper presents two methods to realize it. The first is to associate with each point (x,y) of the wave field a two-dimensional wave number spectrum (k_x,k_y), called local spectrum. The second is to process by other coordinates the wave field along an arbitrary direction, introduced in quantum mechanics for the study of classical billiards, and provided by the Birkhoff variables $(s,\cos\varphi)$. Phase-space diagrams are given by quadratic phase-space distributions. Simulations are presented for wave fields in a 2D planar waveguide for a pedagogical point of view with Gaussian beam or point-source excitation, and nonuniform waveguides as a sudden area expansion chamber and an open billiard with a single incoming mode at the entrance of each of them. In these problems, local spectrum and Birkhoff analysis are used in order to identify the structures hidden in the field. The result is the contribution of different wave vectors which contribute to the field value at the analysis point or at a certain section of the boundary, and show complicated structure of the acoustic field like whispering gallery or diffracted waves. [DOI: 10.1121/1.1898123]

PACS numbers: 43.20.Mv [LLT] Pages: 3478–3488

I. INTRODUCTION

In the present study, we propose methods for local analysis of time-harmonic acoustic wave field in a two dimensional space. It is usual to describe a space-varying signal with complex amplitude $G(\mathbf{r})$ in the wave number domain by means of its wave number spectrum $G(\mathbf{k})$ [Fourier transform of $G(\mathbf{r})$], which may be interpreted as the distribution of the energy of the wave field as a function of the wave number. Nevertheless, it does not give any information about the *local* distribution of the energy as a function of both the space and the wave number. This justifies the introduction of *local* spectrum and the use of phase space $(\mathbf{r},\mathbf{k})$.

The representation of a wave field in the phase space is an old idea which has recently received much attention.[1] This representation implies that the distribution of the energy of the wave field is analyzed simultaneously in both the space- and the wave number domains, as a distribution called local spectrum.[2] Phase-space diagrams are well known in quantum mechanics, where both the position and the momentum describe the behavior of a particle.[3,4] The use of phase-space diagrams in radiation problems was first proposed in optics[2,5,6] and more recently in acoustics,[7,8] where the link between phase-space distribution and ray concept was made.

For propagation in Cartesian coordinates, the wave field in 2D may be described as a function of the four variables (x,y,k_x,k_y), called phase-space distribution.[4] To represent this distribution, the choice proposed in Ref. 4 is to choose an observation point (x,y) and to represent the result as a function of (k_x,k_y). For instance, in a waveguide, Fig. 1(b) shows the ideal local spectrum for an observation point *O* according to the configuration of Fig. 1(a), where three rays (*A*, *B*, and *C*) are supposed to meet at point *O*.

Another point of view was proposed to represent such a function in acoustics, in the framework of underwater acoustics[7,8] to study the propagation in an 2D oceanic waveguide. In these studies, the results are presented as a function of the variables (y,k_y) for a given *x*.

In addition, a method proposed in quantum mechanics[9,10] (with Dirichlet boundary conditions) consists of taking all the information of a wave field in its normal derivatives estimated on the boundary. Phase-space distributions used on such a signal is provided by the so-called Birkhoff variables $(s,\cos\varphi)$, where *s* is the coordinate along the boundary, and $k_{\|}/k=\cos\varphi$ its conjugate normalized *k* wave number parallel to the wall.

The paper is organized as follows. In the following section, we review the most common methods for space wave number representations. This way of representing the wave field is then applied to three problems in acoustics. First, the well-known problem of 2D planar waveguide is revisited in Sec. III in a new point of view. Second, a sudden area expansion is analyzed by this way in Sec. IV. We then consider in Sec. V an open billiard[11] that has been discussed recently. Finally, conclusions are presented in Sec. VI.

II. PHASE-SPACE DISTRIBUTIONS

In the following, we are concerned with the acoustic field $G(\mathbf{r})$ (a time dependence $e^{-j\omega t}$ is assumed and suppressed in the following). The notations in a Cartesian coordinate frame are, for the space coordinates, the vector $\mathbf{r}=(x,y)$, and for the wave number coordinates the vector $\mathbf{k}=(k_x,k_y)$. Expressions such as $\mathbf{k}\cdot\mathbf{r}$ mean the scalar product

 0001-4966/2005/117(6)/3478/11/$22.50

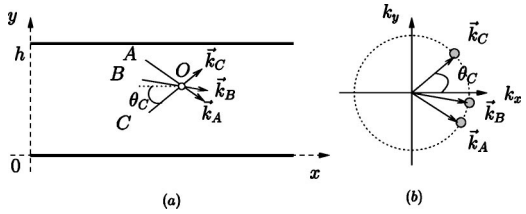


FIG. 1. (a) Spatial behavior in a planar waveguide of three rays *A*, *B*, and *C*, which cross at an analysis point *O*. (b) Ideal local spectrum in *O* showing, in the wave number domain, the contribution of those three rays.

$(k_x x + k_y y)$. Last, expressions like $\int \cdots d\mathbf{r}$ represent $\int\int \cdots dx\, dy$, and all integrations extend from $-\infty$ to $+\infty$ if not specified.

A. Husimi distribution (HD)

The Husimi distribution introduced in quantum mechanics[12] (denoted in the following as HD) is the most widely used tool to analyze the local spectrum of a signal. It is easily shown that the HD is the squared modulus of a Gaussian-windowed Fourier transform (denoted in the following as GWFT) applied in optics[5,6] and acoustics[7,8] or short-space Fourier transform (denoted in the following as SSFT) in signal processing.[13] The HD of a space domain harmonic function $G(\mathbf{r})$ is expressed as

$$HD_G(\mathbf{r},\mathbf{k}) = |\mathrm{GWFT}_G(\mathbf{r},\mathbf{k})|^2, = \left| \int G(\mathbf{r}') w^*(\mathbf{r}-\mathbf{r}') e^{-j\mathbf{k}\cdot\mathbf{r}'} d\mathbf{r}' \right|^2, \quad (1)$$

where GWFT_G denotes the Gaussian-windowed Fourier transform of the function $G(\mathbf{r})$, and $\mathbf{r}$ is the position of the center of the space window $w(\mathbf{r})$, which is supposed to be a Gaussian window. It is a real and positive phase-space distribution.

B. Pseudo-Wigner-Ville distribution (PWVD)

The pseudo-Wigner–Ville distribution (PWVD) of the function $G(\mathbf{r})$ is defined[13,14] by

$$\mathrm{PWVD}_G(\mathbf{r},\mathbf{k}) = \int \left| w\left(\frac{\boldsymbol{\alpha}}{2}\right) \right|^2 G\left(\mathbf{r}+\frac{\boldsymbol{\alpha}}{2}\right) G^*\left(\mathbf{r}-\frac{\boldsymbol{\alpha}}{2}\right) e^{-j\mathbf{k}\cdot\boldsymbol{\alpha}} d\boldsymbol{\alpha}, \quad (2)$$

where $w(\mathbf{r})$ is an even function (called window) which introduces bounds of integration in contrast to the Wigner–Ville distribution[2,5,13,15] (WVD). The PWVD has several significant properties: it is real and not everywhere positive. Moreover, this distribution is quadratic as the HD, rather than linear as the GWFT.[6–8] But, in contrast to HD (except for neighboring terms), it generates large cross terms located midway between components.

C. Modified Wigner–Ville distribution (MWVD)

The modified Wigner–Ville distribution (MWVD) has been introduced by Stankovic[14] *et al.* in the signal-processing framework. It is defined according to

$$\mathrm{MWVD}_G(\mathbf{r},\mathbf{k}) = \frac{1}{\pi^2} \int M(\boldsymbol{\theta}) \mathrm{SSFT}_G(\mathbf{r},\mathbf{k}+\boldsymbol{\theta}) \times \mathrm{SSFT}_G^*(\mathbf{r},\mathbf{k}-\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3)$$

where SSFT is the short-space Fourier transform and $M(\boldsymbol{\theta})$ is a window. Some useful properties may be considered which can be gained by an appropriate choice of the window $M(\boldsymbol{\theta})$. Two cases are

If $M(\boldsymbol{\theta}) = \pi^2 \delta(\boldsymbol{\theta})$, where δ is the Dirac δ function, Eq. (3) leads to $\mathrm{MWVD}(\mathbf{r},\mathbf{k}) = \mathrm{HD}(\mathbf{r},\mathbf{k})$.

If $M(\boldsymbol{\theta}) \equiv 1$, for all $\boldsymbol{\theta}$, Eq. (3) leads to $\mathrm{MWVD}(\mathbf{r},\mathbf{k}) = \mathrm{PWVD}(\mathbf{r},\mathbf{k})$.

This distribution may consequently be seen "between" the HD and the PWVD and it combines the properties of both. It can produce the desired representation of a multicomponent signal such that the distribution of each component $F_q(\mathbf{r})$ is its PWVD, avoiding cross terms between components.

D. Smoothed Pseudo-Wigner–Ville distribution (SPWVD)

It has been shown that smoothing the WVD reduces the importance of amplitude cross terms,[13] but at the expense of smearing the autocomponent concentration as for the PWVD. The SPWVD uses two independent smoothing windows: a spatial averaging window $h(\boldsymbol{\beta})$ and a wave number averaging window $w(\boldsymbol{\alpha})$. This double-smoothing operation produces an improvement of the representation by reducing the cross terms. The SPWVD can be defined as

$$\mathrm{SPWVD}_G(\mathbf{r},\mathbf{k}) = \int \left| w\left(\frac{\boldsymbol{\alpha}}{2}\right) \right|^2 \left[\int h(\boldsymbol{\beta}-\mathbf{r}) G\left(\boldsymbol{\beta}+\frac{\boldsymbol{\alpha}}{2}\right) \times G^*\left(\boldsymbol{\beta}-\frac{\boldsymbol{\alpha}}{2}\right) d\boldsymbol{\beta} \right] e^{-j\mathbf{k}\cdot\boldsymbol{\alpha}} d\boldsymbol{\alpha}. \quad (4)$$

The independent smoothing in the space- and wave number domain yields significant practical advantages, resulting in a great flexibility in the choice of smoothing window, application, and efficient computation.[13]

E. Birkhoff variables

Another phase-space analysis exists according to new coordinates. In quantum mechanics, a natural way of reducing the quantum problem from two dimensions to one is to study the properties of a wave function on the wall which qualifies the field on boundaries. Neumann boundary conditions are required for acoustic problems, and phase-space distribution representation of a wall wave function, also called a Birkhoff map[16] or quantum Poincaré section,[10] is given by phase-space coordinates $(s, \cos\varphi)$. The variables linked to the surface are *s*, an arc length along the boundary, and its conjugate normalized wave number $\cos\varphi = k_{\|}/k$, which is the normalized component of wave number parallel to the wall, where φ is the angle of incidence as shown by Fig. 2. Instead of vectors $\mathbf{r}$ and $\mathbf{k}$, the space and wave num-

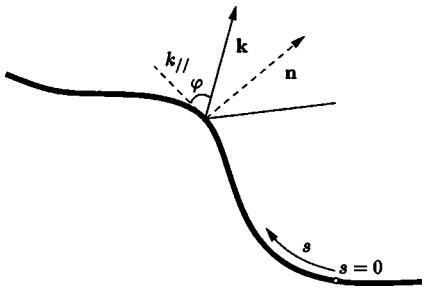


FIG. 2. Birkhoff variables are fixed by specifying the arc length along the boundary parametrized by s and normalized component of wave number parallel to the wall $\cos\varphi = k_{\parallel}/k$.

ber coordinates are, respectively, s and $k_s = k\cos\varphi$. Thus, expression of the scalar product $\mathbf{k}\cdot\mathbf{r}$ means the product $k\cos\varphi \times s$.

III. 2D WAVEGUIDE PROBLEM

In this section we propose phase-space distributions in various settings. The concepts and the qualitative observations discussed in the Introduction and in Sec. II are applied to a 2D planar waveguide with Gaussian beam or a point-source excitations.[17–19] The purpose of choosing a Gaussian beam or a point source is to provide a convenient test case to compare numerical results. The phase-space distribution for a Gaussian beam or a point-source excitation in a planar waveguide is formally obtained by substituting either the entire or a part of the modal Green's function, which is taken as a reference solution, into the definitions of the HD, the PWVD, the MWVD, and the SPWVD.

A. Gaussian beam excitation of a 2D waveguide

Because of the Neumann boundary conditions at $y=0$ and $y=h$ of Fig. 3(a), the modes are discrete with a vertical wave number $k_m = m\pi/h$, with m the mode index and h the height of the planar waveguide. A propagating mode can thus be regarded as two interfering sets of rays at angles, $\varphi_m = \arcsin(k_m/k)$, and this interpretation has been widely used.[20] A cluster of modes is known to produce an interference maximum along a trajectory (emanating from the source) equivalent to the path of the "modal ray" φ_m for the central mode in the group.[17,18] The field of a modal Gaussian beam may be expressed as

$$\sum_{m=M_1}^{M2} e^{-[(m-m_0)^2]/2\sigma^2} G_m(\mathbf{r},\mathbf{r}_0), \tag{5}$$

where $e^{-[(m-m_0)^2]/2\sigma^2}$ is a Gaussian window, with m_0 the index of the central mode of the mode bundle, M_1 and M_2 the mode indexes which determine bounds of the mode bundle, and G_m the modal Green function with mode index m, given by

$$G_m(\mathbf{r};\mathbf{r}_0) = \left(\frac{2-\delta_{m0}}{2h}\right) j \frac{\cos\left(\frac{m\pi}{h}y_0\right)}{\sqrt{k^2-\left(\frac{m\pi}{h}\right)^2}} \times \cos\left(\frac{m\pi}{h}y\right) e^{j\sqrt{k^2-(m\pi/h)^2}|x-x_0|}, \tag{6}$$

with δ_{m0} the Kronecker tensor and $k=\omega/c_0$. Note that the Green function is $G=\Sigma_{m=0}^{\infty} G_m$.

1. Phase space of four variables

For a 2D waveguide, the parameters used for the computations are the height $h=49.97\lambda$, which allows 100 propagative modes where λ is the wavelength of the source, the wave velocity is $c_0 = 340\,\mathrm{m\cdot s^{-1}}$, and the density is $\rho_0 = 1.2\,\mathrm{Kg\cdot m^{-3}}$. The waist of the Gaussian beam has been arbitrarily located at $x_0=0$ and $y_0=h$, and the choice of the central mode in the group is $m_0=69$. We denote the mode bundle width as follows: ΔM and take the value $\Delta M=25$ according to criteria given in Ref. 18. It is weighted by a Gaussian window[21] with a beamwidth parameter as $\alpha=0.04$. Phase-space distributions are computed at the point $(x,y) = (75\lambda, 20\lambda)$. Here and in the later sections, simulations were realized with a 512-point FFT algorithm using zero padding [spatial window $w(\mathbf{r})$, which is used in both the HD and the PWVD, is such that its width along x- and y axes is $W_x = W_y = 8\lambda$]. Results given in Fig. 3(b) and Fig. 3(c) show, respectively, the HD and the PWVD at the observer point. The cross, which is located on the radiation circle, describes a geometric ray emanating from the point source, which reaches the observer point with one reflection at the lower boundary. Phase-space distributions are, at sufficiently high frequency, highly localized in phase space around the "skeletal point"[7,8] with an exponential decay [Gaussian in fact, due to spatial window $w(\mathbf{r})$]. Considering simulation shown in Fig. 3(b) and Fig. 3(c), the PWVD offers higher resolution than the HD for a monocomponent signal. Simply stated, the PWVD uses finite bounds of integration in contrast with the WVD. This results in a function which, relative to the true WVD, is smoothed with respect to the frequency domain only, whereas the HD is the result of the true WVD smoothed in both frequency and space domain.[5,12,13] The large positive values, named "islands,"[10] which are located on the radiation circle, are well suited according to the departure angle of the Gaussian beam φ_{69}. Indeed, as discussed in earlier sections, the HD is equivalent to a blurred PWVD.

2. Birkhoff analysis

The horizontal axis of the Birkhoff map s, which describes the position along the boundary, varies from 0 to 200λ with λ the wavelength of the source [Fig. 3(a)]. Only one variable (noted x) is necessary for such a representation in the phase space. The coordinate y is arbitrarily chosen such that $y=0$. The Gaussian beam hits the lower boundary along the parameter s of the 2D waveguide at two different fixed points, noted $s=s_1$ and $s=s_2$, and bounces off with

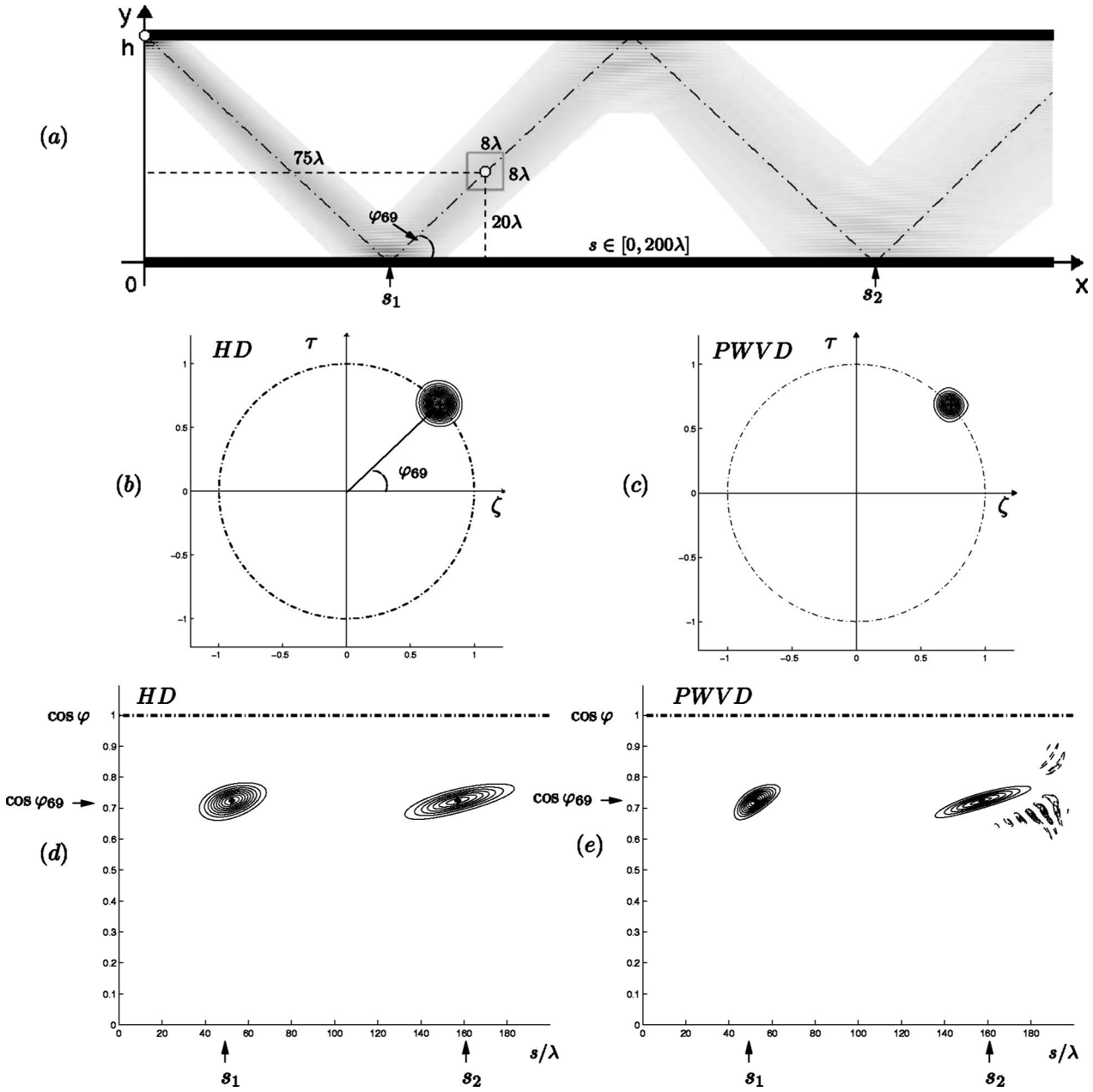


FIG. 3. (a) Magnitude of the Gaussian beam in the 2D waveguide. Local spectrum ($\zeta=\cos\varphi$, $\tau=\sin\varphi$) normalized by the wave number k for the HD (b) and the PWVD (c) at the observer point with Cartesian coordinates $(x,y)=(75\lambda,20\lambda)$ for a Gaussian beam excitation at the upper interface with departure angle corresponding to φ_{69}. Dashed circle (—■) represents the normalized radiation circle, which is the trigonometric circle. Birkhoff map: HD (d) and PWVD (e). Skeletal points are represented by crosses.

the same angle α as shown in Fig. 3(a). Simulations were performed, here and in the later sections, with a 2048-point FFT algorithm. Spatial window $w(\mathbf{r})$ used in the HD [Fig. 3(d)], and in the PWVD [Fig. 3(e)], is applied such as its width along s axis is $W_s=51\lambda$. Figure 3 shows the behavior of a Gaussian beam in a 2D waveguide according to the Birkhoff map with the help of the HD [Fig. 3(d)] and the PWVD [Fig. 3(e)]. Islands are centered on the coordinates of the ray path of the central group of the mode, which is represented by skeletal points (crosses). Those skeletal points describe the first and second impact on the lower interface. Figure 3(e) shows that the PWVD provides increased resolution relative to the HD, but it generates interference terms.

B. Line source excitation of a 2D waveguide

With a line source excitation at $\mathbf{r}_0=(x_0,y_0)$, the solution is the 2D Green function given by

$$G(\mathbf{r},\mathbf{r}_0)=\sum_{m=0}^{\infty} G_m(\mathbf{r},\mathbf{r}_0), \tag{7}$$

where $G_m(\mathbf{r},\mathbf{r}_0)$ has been defined in Eq. (6).

1. Phase space of four variables

As for the Gaussian beam solution, the duct height is $h=49.97\lambda$, which allows 100 propagative modes. The point source is on the upper boundary at $(x_0=0,y_0=h)$. The guided mode series in Eq. (7) is truncated at the order

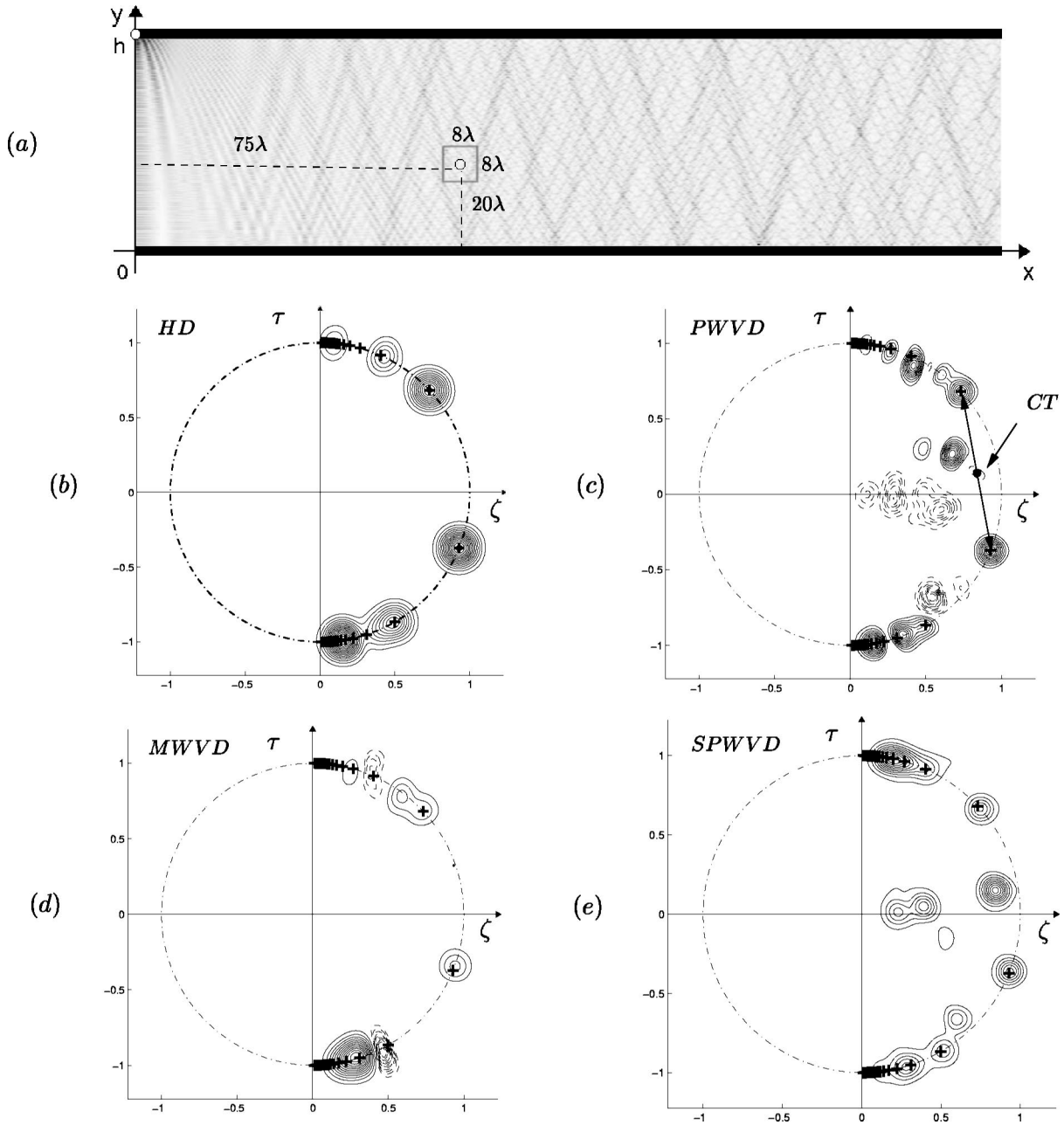


FIG. 4. (a) Magnitude of the pressure field in the 2D waveguide. Local spectrum ($\zeta=\cos\varphi$, $\tau=\sin\varphi$) normalized by the wave number k for the HD (b), the PWVD (c), the MWVD (d), and the SPWD (e) at the observer point with Cartesian coordinates $(x,y)=(75\lambda,20\lambda)$ with a point source at the upper interface $(x_0,y_0)=(0,h)$. Positive and negative contour plot are, respectively, represented by solid line and dashed line. Dashed circle (—■) represents the normalized radiation circle, which is the trigonometric circle.

$m=150$, which means that there are 50 evanescent modes. We now consider multicomponent signal (several islands) with cross terms which correspond to interference between islands. The cross terms may have a peak value higher than the auto components.[13] In this section, we investigate the space–wave number resolution of representations in a 2D waveguide given by the HD, the PWVD, and the other two distributions which have qualities in cross-terms reduction, the MWVD and the SPWVD. Phase-space distributions are computed at the point $(x,y)=(75\lambda,20\lambda)$. Results given in Fig. 4 show the four phase-space distributions at the observer point. Crosses which are located on the radiation circle describe geometric rays emanating from the point source, which reach the observer point and represent, respectively, the direct and reflected ray families. A better accuracy is obtained to localize local wave number with the PWVD [Fig. 4(c)] than the other phase-space distributions. Nevertheless, it is very difficult to interpret the PWVD because of some interference terms[14] and negative values.[22] Islands are concentrated on the radiation circle, while interferences are centered between the corresponding islands.[14] The resulting interference forms a type of noise that may obscure a weaker, nearby component. An example is presented in Fig. 4(c) by the letters *CT*. In contrast, the HD [Fig. 4(b)] does not tend to have the ghost structures mentioned above for the PWVD, except for neighboring terms.[13] One desirable property of the HD is that $\mathrm{HD}(\mathbf{r},\mathbf{k})\geqslant 0$ and the interpretation of features in the HD tends to be easier; on the other hand, the features

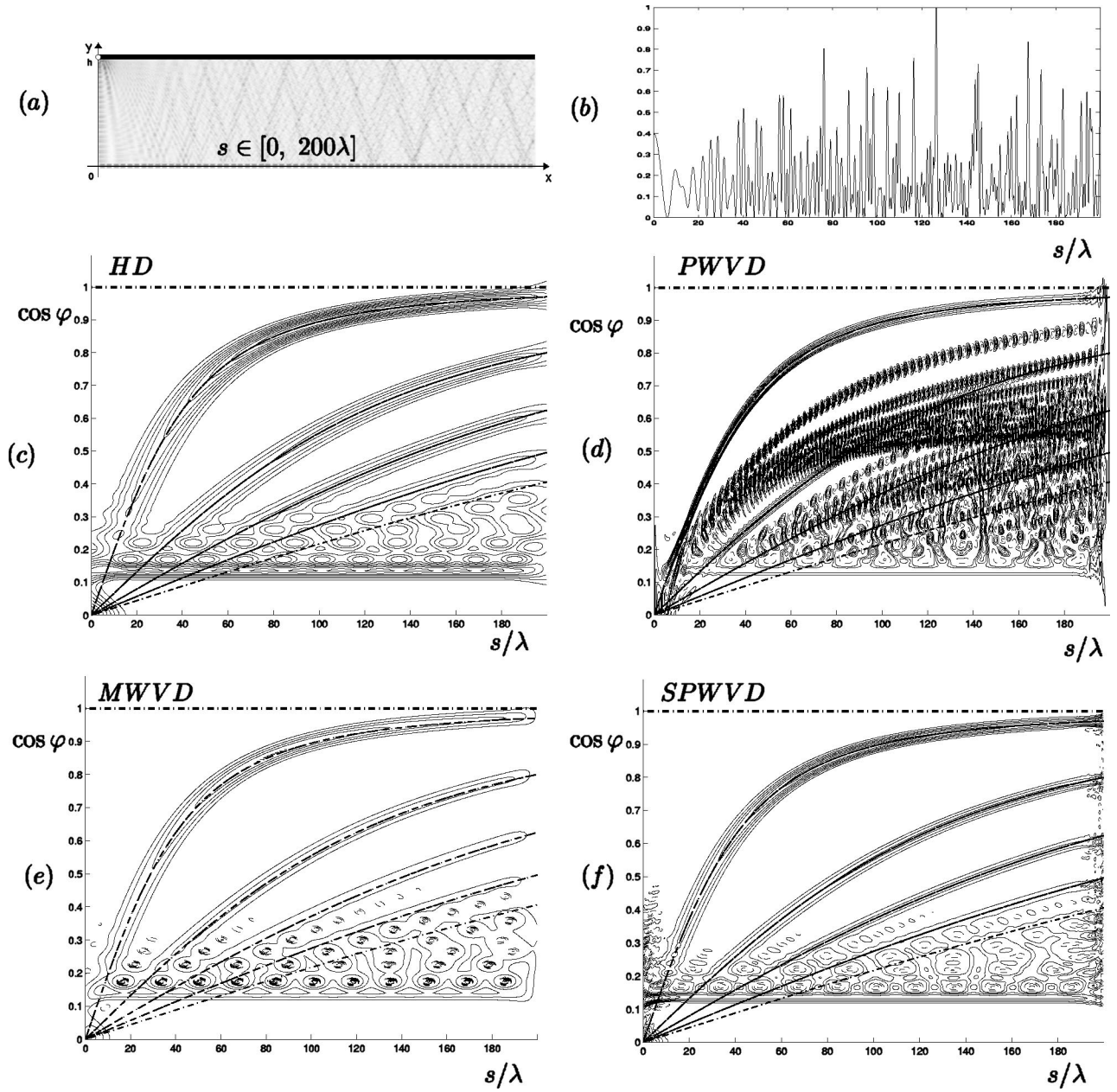


FIG. 5. Phase-space analysis according to Birkhoff variables $(s,\cos\varphi)$. The duct height is $h=49.97\lambda$ and a point source is located at the Cartesian coordinates $(x_0,y_0)=(0,h)$ (a). (b) Normalized magnitude of the pressure field on the lower interface. (c) the HD; (d) the PWVD; (e) the MWVD; and in (f) the SPWVD. Skeletal lines are represented by dashed line.

tend to show rather poor wave number resolution. Many other distributions have been developed with the purpose of cross-term reduction. In contrast to the PWVD, the MWVD [Fig. 4(d)] is equal to the sum of the PWVD of the individual components, only if auto terms are far away from each other. It confirms the absence of ghost terms inside the radiation circle compared to the PWVD. A substantial reduction of the PWVD cross terms can be obtained by spatial and wave number smoothing. The SPWVD [Fig. 4(e)] realizes this filtering and gives islands with good resolution. Some cross terms are still present but less than the MWVD and the PWVD. It is clear, therefore, that the HD [Fig. 4(b)] and the SPWVD [Fig. 4(e)] exhibit all essential structures necessary to obtain meaningful information, whereas the PWVD [Fig. 4(c)] and the MWVD [Fig. 4(d)] contain extremely complex patterns. A post-treatment could nevertheless be considered to find auto terms with the help of cross terms.

2. Birkhoff analysis

Let us take the same properties as in the above section. The problem of surface wave field in a 2D waveguide is essentially that of finding beam fields created by a point source located on the upper wall $(x_0=0,y_0=h)$ into the $x>0$ half plane, for $y=0$. In the Birkhoff map, the information is centered around skeletal lines that coincide with the location of the geometrical rays emanating from the $x_0=0$ plane at $y_0=h$, and passing through observer point along the lower interface of the 2D waveguide $s\in[0,200\lambda]$ with λ the wavelength of the source [Fig. 5(a)]. The skeleton which

corresponds to the direct and reflected rays, represented by dashed line, and the phase-space distribution are shown in Fig. 5. All the windows used in simulations are Gaussian. The HD [Fig. 5(c)] and the PWVD [Fig. 5(d)] give results in good agreement with the multibranch skeleton. In accordance with previous conclusions about the PWVD, we obtain a better resolution than the HD, but also a lot of undesirable cross terms, which clearly prevents any straightforward reading. Except for the direct ray, which may be easily identified, the skeletal lines become closer and interfere strongly for high-order reflected ray and can no longer be separated. Distributions such as the MWVD [Fig. 5(e)] and the SPWVD [Fig. 5(f)] have been developed to reduce the cross-term components. Concerning the SPWVD, the use of a double-smoothing operation will produce a further improvement in the cross-term reduction. We observe that SPWVD performs better than the other three distributions according to an increased computational effort.

3. Discussion

In order to identify more complicated structure hidden in a wave field, as we propose in the next sections, and for easier use and interpretation of results, the HD will be used for the next (x,y,k_x,k_y) representation and Birkhoff map.

C. Inverse problem

In direct radiation problems, the source term, the radiative properties of medium, and the boundary conditions are given. The acoustic field is to be determined with those data. On the other hand, in inverse radiation problems, either the radiative properties, or the source terms, or the boundary conditions are to be determined from the knowledge of the measured radiation data. In this inverse analysis, $\Delta x=(x-x_0)$, the propagation distance between the observer point and the source point placed on the upper interface, and h, the height of the 2D waveguide, are regarded as unknown, but other quantities are known as the measured incident radiation field on the lower boundary according to conditions given in Sec. III B 1. Simulation given in Fig. 5(d) shows, with the help of the PWVD, the best image resolution for the direct radiation. This one is described by a skeletal line given by the simultaneous solution of

$$\begin{cases} \Delta x \tan\varphi = \pm h, \\ k_x = k\cos\varphi, \end{cases} \tag{8}$$

with h the unknown height of the 2D waveguide when the straight ray trajectory leaves the unknown initial plane x_0 at the upper interface and φ the incident angle, corresponding to the direct ray. It is easy to find Δx and h in fixing a value of the wave number parallel to the wall $k_x=k\cos\varphi$. The direct ray skeletal line given by the PWVD, which corresponds to the chosen k_x, gives the position $x=s$ according to the Birkhoff map. Then, the equations given by Eq. (8) can furnish the value of h. Different works in inverse radiation problems could be investigated with the help of those tools.

IV. SUDDEN AREA EXPANSION PROBLEM

In this section, HD will be applied to both the local spectrum and the Birkhoff analysis in a sudden area expansion problem, and their results will be compared. The general duct system to be analyzed consists of two semi-infinite pipes (numbered with indexes $q=1, 2$) of heights h_1 and h_2, respectively, which are joined together at $x=0$ [Fig. 6(a)]. Note that rigid duct wall boundary conditions have been assumed. An incident mode, which can be expressed as two interfering plane waves, propagates from left to right towards the expansion, where it is partly reflected and partly transmitted into the larger channel. The method to calculate the behavior of an incident mode as it propagates past a single discontinuity has been outlined previously and is discussed briefly here. The acoustic pressure, which is the homogeneous problem (without sources) inside a sudden area expansion, can be expressed using infinite series as

$$p(\mathbf{r})=\sum_n \psi_n(y)P_n(x), \tag{9}$$

where ψ_n are the eigenfunctions, which are expressed as

$$\psi_{qn}(y)=\sqrt{\frac{2-\delta_{n0}}{h_q}}\cos\left(\frac{n\pi}{h_q}y\right) \quad \text{with} \quad q=1,2. \tag{10}$$

They are the classical transverse modes in a straight duct.

$$p_1(\mathbf{r})=p^{(i)}(\mathbf{r})+p^{(r)}(\mathbf{r}), \tag{11}$$

represents the acoustic field in the smaller channel. The incident and the reflected pressure field are, respectively, denoted by $p^{(i)}$ and $p^{(r)}$.

$$p_2(\mathbf{r})=p^{(t)}(\mathbf{r}), \tag{12}$$

is the transmitted pressure field in the larger channel. Boundary conditions at $x=0$, for pressure and volume velocity, leads to the expressions

$$p_1(\mathbf{r})=p_2(\mathbf{r}), \quad \text{with} \quad x=0 \quad \text{and} \quad \forall y\in[0,h_1], \tag{13}$$

and

$$\begin{cases} \partial_x p_1(\mathbf{r})=\partial_x p_2(\mathbf{r}), & \text{with } x=0 \text{ and } \forall y\in[0,h_1], \\ \partial_x p_2(\mathbf{r})=0, & \text{with } x=0 \text{ and } \forall y\in[h_1,h_2]. \end{cases} \tag{14}$$

The reflection and transmission coefficients are calculated by the modal decomposition approach[23] and are given in the Appendix. The problem is investigated by local phase-space analysis in different areas of such a waveguide.

A. Phase space of four variables

Consider the channel system illustrated in Fig. 6. Two semi-infinite channels, of height $h_1=35.33\lambda$ and $h_2=49.62\lambda$, with λ the wavelength of the frequency regime, join together at $x=0$. In this case, we consider frequencies that allow 70 and 100 propagating modes, respectively, for the smaller and the larger channels. The present example illustrates the behavior of an incoming mode with index $n=19$. After truncating a sufficient number of modes ($N=150$), the reflected and transmitted pressure fields are ob-

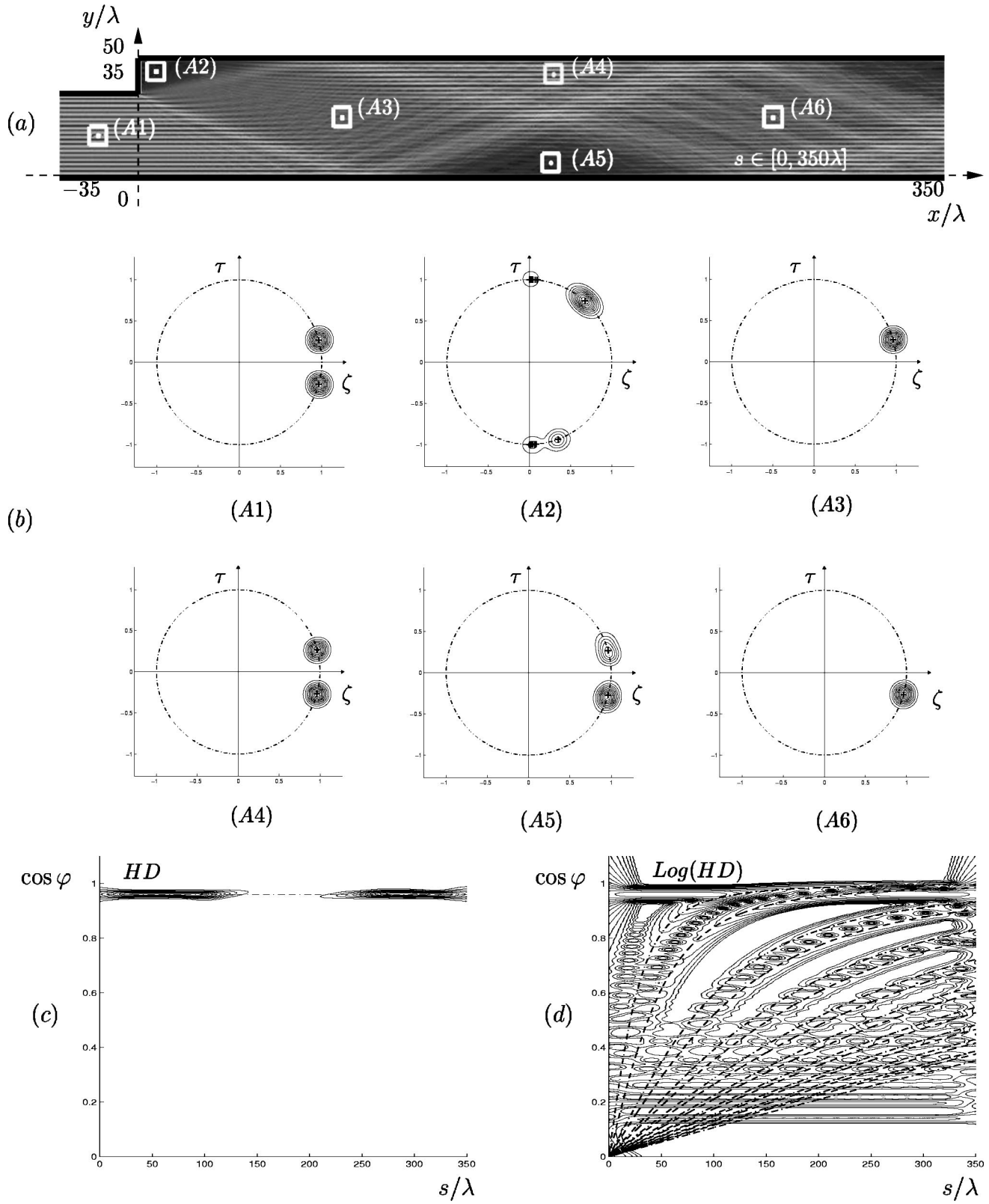


FIG. 6. (a) Magnitude of the pressure field in a sudden area expansion with an incoming mode (from the left) number 19. Different points are plotted where local spectrum are realized. (b) Phase-space analysis with the help of the HD for points $(A1)=(-18\lambda,-18\lambda)$, $(A2)=(8\lambda,44\lambda)$, $(A3)=(88\lambda,25\lambda)$, $(A4)=(180\lambda,43\lambda)$, $(A5)=(179\lambda,6\lambda)$, and $(A6)=(276\lambda,25\lambda)$. (c) and (d) the HD and the log(HD) with Birkhoff variables, respectively. Skeletal lines are represented by dashed line.

tained with the modal decomposition method. We can then obtain a distribution of outgoing modes for each incoming mode. Figure 6(a) shows the spatial behavior of the acoustic pressure field in such a waveguide for the incoming mode 19. We can distinguish different areas in the larger channel, when a local analysis may be carried out. Local spectrum are estimated by HD at points $(A1)$, $(A2)$, $(A3)$, $(A4)$, $(A5)$, and $(A6)$. The widths of the spatial window $w(\mathbf{r})$ are W_x

$=W_y=8\lambda$. The first point $(A1)$ shows the contribution of incoming mode number 19. This incoming mode is uniformly distributed as two plane waves in opposite directions which are translated in the local spectrum by two maxima on the right-hand side of the radiation circle (i.e., $\zeta>0$, corresponding to propagation from left to right). This figure confirms that incoming mode number 19 is practically not reflected because no island is presented in the left side of the radiation circle. Points labeled $(A3)$, $(A4)$, and $(A6)$ show the predominance of outgoing mode number 27 in the larger channel, which is the predominantly transmitted outgoing mode. First, local spectrum $(A3)$ illustrates, as for $(A1)$, the contribution of mode number 27 on the radiation circle. This is furthermore proved by the location of skeletal points of mode 27 represented by crosses, which are in accordance with positions of islands of mode number 27. Moreover, interesting features are local spectra $(A3)$ and $(A6)$, which illustrate, respectively, upward and downward mode components. Then, for the shadow area $(A2)$ clearly present diffracted waves created by the corner at position $(0,h_1)$. Such diffractive points can be considered as new wave sources whose strength is proportional to the strength of the incident wave. Skeletal points are in accordance with islands given by the HD. They describe geometric rays emanating from the corner which reach the analysis point $(A2)$ and represent, respectively, the direct and reflected ray families as seen before. Last, for the point $(A5)$, no definite conclusion may be given. Birkhoff analysis, as presented in next section, gives more information about the area around this point, in order to find if it corresponds to a mode behavior or to a geometrical behavior.

B. Birkhoff analysis

In order to describe the behavior of the wave field around the point $(A5)$, s is arbitrarily chosen along the lower interface, such that $y=0$ and x varies from 0 to 350λ. Figure 6(c) shows the result of Birkhoff analysis according to the HD. First, two islands located at the beginning and the end of segment s illustrate the contribution of transmitted mode number 27. Note that a skeletal horizontal line representing the longitudinal wave number component of mode number 27 is in accordance with the position of these two islands. Between both islands, a shadow area in the Birkhoff map seems to prove that around the point $(A5)$ the behavior of the wave field is not a mode one. The stellar representation[24] is displayed in Fig. 6(d). This simple representation consists of ploting the HD with a logarithmic (instead of linear) scale. Both representations [HD and log(HD)] provide complementary viewpoints. The last one encodes the fully wave function and shows the contribution of diffracted waves created by the corner at position $(0,h_1)$, in addition to the predominance of mode number 27. Skeletal lines describe geometric rays emanating from that point and represent the direct and reflected ray families as seen in Sec. III B 2. In conclusion, the shadow area around the point $(A5)$ is generated by diffracted waves from the corner of the sudden area chamber, and such an interpretation is provided by the help of the stellar representation.

V. OPEN BILLIARD

Finally, HD will be applied to both the local spectrum and the Birkhoff analysis in an open billiard. It consists of analyzing, with those two tools, the behavior of a wave field in a waveguide composed by a resonator, coupled to two ducts[12] as shown in Fig. 7(a). The resonator we consider here has the shape of half an annular billiard.[25] It consists of a small disk of radius $R2$ located inside a larger disk of radius $R1$ with a displacement δ of the disk centers. In order to solve the wave problem numerically by using a finite elements method, incoming and outgoing wave functions at the entrance of the two ducts, say left (L) and right (R) ducts, are given by

$$p_L(x,y)=e^{jky}+Re^{-jky}, \tag{15}$$

and

$$p_R(x,y)=Te^{-jky}, \tag{16}$$

where the notations R and T correspond, respectively, to the reflection and transmitted coefficients, and where k is the wave number of the propagating plane wave in the y direction. Neumann boundary conditions are furthermore required on the walls.

Numerical results

The left and right ducts of the billiard of Fig. 7(a) have identical width $h_L=h_R=6.05\lambda$, with λ the wavelength of the frequency regime. The upper part of the waveguide is made up of the larger disk with radius $R1=18.14\lambda$ and the small disk with radius $R2=3.02\lambda$. The displacement of the disk centers is about $\delta=9.07\lambda$. We consider a frequency that allow 13 propagating modes in the left and right ducts. With the help of the finite element method, the acoustic pressure field in the spatial domain is obtained and shown in Fig. 7(a). Local spectra are then estimated by HD at points $B1$ and $B2$ with a spatial window of widths $W_x=W_y=6\lambda$. At the same time, we choose two Birkhoff analyses along curvilinear coordinates $s1$ and $s2$, as shown in Fig. 7(a). The first Birkhoff analysis along $s1$ is presented in Fig. 7(c). Trajectories close to the convex boundary of the waveguide are defined as a whispering gallery (WG). These trajectories take up the major part of the Birkhoff map, as seen in Fig. 7(b). Two kinds of WG may be distinguished, with, respectively, order $q=2$ and $q=4$ corresponding to the number of rebounds on the $s1$ segment. In order to supplement these results, we consider the ray tracing inside the same waveguide. Boundaries act as mirrors for rays in accordance with the laws of geometrical acoustics. We can see in the inset of Fig. 7(b) ray tracing of WG of order 2 and 4, respectively, noted $WG2$ and $WG4$. In Fig. 7(d), the second Birkhoff analysis gives information in the shadow area, where non-WG trajectories are trapped inside the cavity due to creeping and diffracted waves. Moreover, the upper part of Fig. 7(d) (i.e., $\cos\varphi>0$) corresponds to the forward propagation, while its lower part (i.e., $\cos\varphi<0$) corresponds to the backward propagation. A predominant island on the upper left part of the Birkhoff map illustrates the presence of creeping waves labeled in Fig. 7(b) as CW. Figure 7(b) shows ray tracing of that kind of creeping ray. Moreover, a backward creeping waves contribution

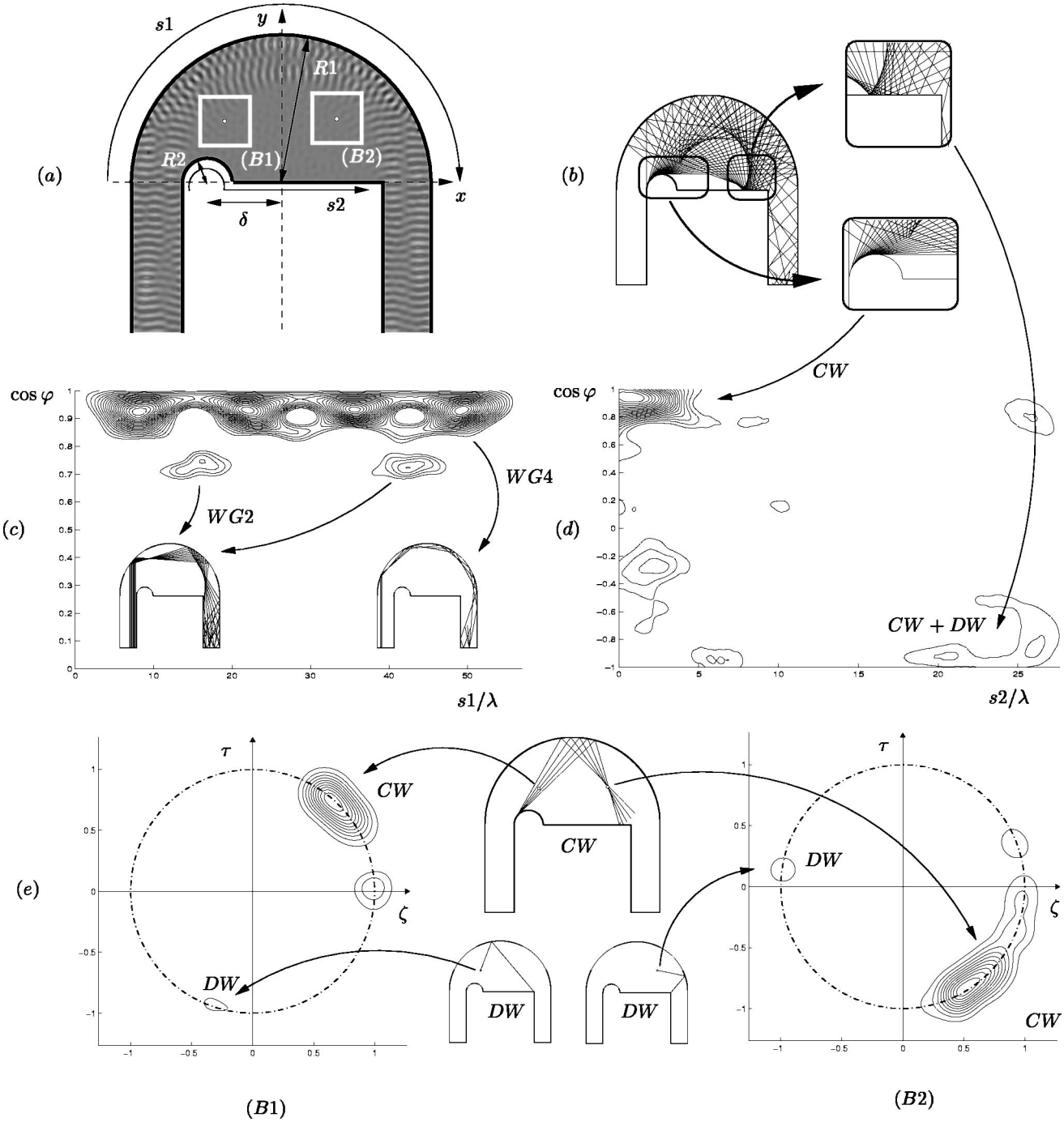


FIG. 7. (a) Magnitude of the pressure field in the space domain. (b): Typical creeping rays. (c) Birkhoff map with the HD along the curvilinear segment $s1$. (d) Birkhoff map with the HD along the curvilinear segment $s2$. (e) Local spectrum with the HD for points $(B1)=(-6.95\lambda,7.56\lambda)$ and $(B2)=(7.25\lambda,7.86\lambda)$.

is shown with a small part of the diffracted waves from the corner placed at points $(12.09\lambda,0)$, in the lower right part of Fig. 7(d). Concerning local spectra in $(B1)$ and $(B2)$, we confirm the presence of creeping waves in the shadow area represented by two large islands noted CW in Fig. 7(e), and is confirmed by ray tracing of a pencil of creeping waves shown in the inset of Fig. 7(e). Last, small islands, noted DW, are located on the left part of radiation circle. They show the role of diffracted waves in the shadow area, leaving the diffractive point in the wedge of the waveguide at $(12.09\lambda,0)$ as shown in the inset of Fig. 7(e). Nevertheless, there is additional structure in the HD at point $(B1)$ and $(B2)$ which has no simple explanation.

VI. CONCLUSION

In this paper, the importance of phase space in investigating wave phenomena has been presented with some applications in waveguide. Included are different phase-space distributions such as the HD, the PWVD, the MWVD, and the SPWVD. Two kinds of phase-space analysis are pre-

sented with the Birkhoff map and the four-variable phase-space domain. Information about the structure of the propagating field can be gained by examining its phase-space distribution. This distribution is centered around areas, in the Birkhoff map or in the local spectrum, that provide the local preferred radiation direction of the field. The localization properties have been schematized in terms of regions surrounding geometric radiation or diffracted radiation in the case of the sudden area expansion chamber, or by creeping waves and whispering gallery behavior in Sec. V.

APPENDIX: REFLECTION AND TRANSMISSION COEFFICIENT VIA MULTIMODAL APPROACH

In this appendix we give the expressions for the reflection and transmitted coefficient presented in Sec. IV. According to matricial terminology, Eq. (9) takes the form of $p = {}^t\boldsymbol{\psi}\mathbf{P}$, where $\mathbf{P}=(P_n)_{n\geqslant 0}$ and $\psi=(\psi_n)_{n\geqslant 0}$ are column vectors with scalar coefficients P_n and ψ_n, respectively. The modal decomposition[23] transforms Eqs. (13) and (14), with the help of Eqs. (11) and (12), and can be expressed for $x=0$ and $\forall y\in[0,h_1]$ as

$$\begin{cases} F\mathbf{P}^{(t)}-\mathbf{P}^{(r)}=\mathbf{P}^{(i)}, \\ K_2\mathbf{P}^{(t)}+{}^tFK_1\mathbf{P}^{(r)}={}^tFK_1\mathbf{P}^{(i)}, \end{cases} \tag{A1}$$

where K_i and F are matrices given by

$$K_{imn}=\begin{cases} \sqrt{k^2-(\alpha_n^{(i)})^2}, & \text{for propagative mode,} \\ -j\sqrt{(\alpha_n^{(i)})^2-k^2}, & \text{for evanescent mode,} \end{cases} \tag{A2}$$

where the associated eigenvalues are $\alpha_n^{(i)}=n\pi/h_i$, which obey the eigenproblem $\Delta_\perp\psi_n=-\alpha_n^2\psi_n$ with ψ_n eigenfunctions.

The matrix F is calculated via

$$F_{mn}=\int_0^{h_1}\psi_{1m}(y)\psi_{2n}(y)dy, \tag{A3}$$

and takes the form of

$$F_{mn}=\begin{cases} \sqrt{a}, & \text{if } m=n=0, \\ 0, & \text{if } n=0 \text{ and } m>0, \\ \dfrac{\sqrt{2a}}{\pi na}\sin(n\pi a), & \text{if } m=0 \text{ and } n>0, \\ \dfrac{\sqrt{a}}{2\pi na}(-1)^m\sin(n\pi a), & \text{if } m=na, \\ \dfrac{2\sqrt{a}}{\pi na}(-1)^m\left(\dfrac{(na)^2}{(na)^2-m^2}\right)\sin(n\pi a), & \text{otherwise,} \end{cases} \tag{A4}$$

with $a=h_1/h_2$.

For an incident mode of order N, the pressure column vector is

$$\begin{cases} \mathbf{P}^{(i)}=e^{j\sqrt{k^2+(n\pi/h_1)^2}x}, & \text{for } n=N, \\ \mathbf{P}^{(i)}=0, & \text{otherwise.} \end{cases} \tag{A5}$$

With this condition, the resolution of system (A1) gives the expression of the reflected and transmitted column vector pressure field $\mathbf{P}^{(r)}=R\mathbf{P}^{(i)}$ and $\mathbf{P}^{(t)}=T\mathbf{P}^{(i)}$, with the reflection and transmitted coefficient matrices given by

$$\begin{cases} R=-({}^tFK_1)^{-1}(K_2-{}^tFK_1F)(K_2+{}^tFK_1F)^{-1}\,{}^tFK_1, \\ T=2(K_2+{}^tFK_1F)^{-1}\,{}^tFK_1. \end{cases} \tag{A6}$$

[1] See the special issue, J. Opt. Soc. Am. A **17**(12), 2274–2542 (2000).
[2] M. J. Bastiaans, in "Transformations in optical signal processing," edited by W. T. Rodes, J. R. Fienup, and B. E. A. Saleh, Proc. SPIE **373**(2), 49 (1981).
[3] S. W. McDonald, "Phase-space eikonal method for treating wave equations," Phys. Rev. Lett. **54**(12), 1211–1214 (1985).
[4] H. U. Baranger, D. P. DiVincenzo, R. A. Jalabert, and D. Stone, "Classical and quantum ballistic-transport anomalies in microjunctions," Phys. Rev. B **44**(19), 10637–10675 (1991).
[5] M. A. Alonso and G. W. Forbes, "Phase-space distributions for high-frequency fields," J. Opt. Soc. Am. A **17**(12), 2288–2300 (2000).
[6] B. Z. Steinberg, E. Heyman, and L. B. Felsen, "Phase space beam summation for time harmonic radiation from large apertures," J. Opt. Soc. Am. A **8**(1), 41–59 (1991).
[7] B. Z. Steinberg and J. J. McCoy, "Marching acoustic fields in a phase space," J. Acoust. Soc. Am. **93**(1), 188–204 (1993).
[8] B. Z. Steinberg and R. Birman, "Phase space marching algorithm in the presence of a planar wave velocity discontinuity. A qualitative study," J. Acoust. Soc. Am. **98**(1), 484–494 (1995).
[9] P. Bellomo and T. Uzer, "Quantum scars and classical ghosts," Phys. Rev. A **51**(2), 1669–1672 (1995).
[10] B. Crespi, G. Perez, and S.-J. Chang, "Quantum Poincaré sections for two-dimensional billiards," Phys. Rev. E **47**(2), 986–991 (1993).
[11] R. G. Naztmitdinov, K. N. Pichugin, I. Rotter, and P. Šeba, "Whispering gallery modes in open quantum billiards," Phys. Rev. E **64**, 056214 (2001).
[12] K. Husimi, "Some formal properties of the density matrix," Proc. Phys. Math. Soc. Jpn. **22**, 264 (1940).
[13] L. Cohen, *Time Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[14] S. Stanković, L. Stanković, and Z. Uskoković, "On the local frequency group shift and cross terms in some multidimensional time-frequency analysis," IEEE Trans. Signal Process. **43**(7), 1719–1724 (1995).
[15] E. Wigner, "On the quantum correction for thermodynamic equilibrium," Phys. Rev. **40**, 749–759 (1932).
[16] A. M. Ozorio de Almedia and R. O. Vallejos, "Poincaré's recurrence theorem and the unitarity of the S-matrix," Chaos, Solitons Fractals **11**, 1015–1020 (2000).
[17] K. M. Guthrie and C. T. Tindle, "Ray effects in the normal mode approach to underwater acoustics," J. Sound Vib. **47**(3), 292–295 (1976).
[18] A. Kamel and L. B. Felsen, "On the ray equivalent of a group of modes," J. Acoust. Soc. Am. **71**(6), 1445–1452 (1982).
[19] A. Kamel and L. B. Felsen, "Hybrid ray-mode formulation of parallel plane waveguide Green's functions," IEEE Trans. Antennas Propag. **AP-29**(4), 637–649 (1981).
[20] P. Morse and K. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, 1986).
[21] I. T. Lu and L. B. Felsen, "Observables due to beam-to-mode conversion of a high-frequency Gaussian P-wave input in a aluminum plate vacuum," J. Acoust. Soc. Am. **87**(1), 42–53 (1990).
[22] D. Dragoman, "The origin of negative values of the Wigner distribution function," Optik **111**(4), 179–183 (2000).
[23] V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross section waveguides by modal decomposition," J. Acoust. Soc. Am. **100**, 2034–2048 (1996).
[24] J. M. Tualle and A. Voros, "Normal modes of billiards portrayed in the stellar (or nodal) representation," Chaos, Solitons Fractals **5**(7), 1085–1102 (1995).
[25] S. D. Frischat and E. Doron, "Dynamical tunneling in mixed systems," Phys. Rev. E **57**(2), 1421–1443 (1998).

Low-frequency wind noise correlation in microphone arrays[a)]

F. Douglas Shields[b)]
Miltec Research and Technology, Coliseum Drive, University, Mississippi 38677

(Received 2 July 2004; revised 2 February 2005; accepted 2 February 2005)

A three-axis orthogonal microphone array with ten sensors in each arm has been used to study wind noise in the frequency range from 0.05 to 50 Hz. Simultaneous measurements were made of the three components of the varying wind velocity. Measurements have been made for wind speeds from 4 to 7 m/s at three different sites. The frequency-dependent correlation of the wind noise over a range of wind velocities and atmospheric and environmental conditions in the downwind direction varies as $\exp(-3.2X)\cos(2\pi X)$. For the crosswind and vertical directions, the correlation decays approximately as $\exp(-7Y)$, where X is the separation in wavelengths in the downwind direction and Y is this separation in the crosswind or vertical direction. Over a limited range of wave numbers, the power density spectra of the varying wind velocity varied as the wave number to the $-\frac{5}{3}$ power and the pressure spectra as the $-\frac{7}{3}$ power. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1879252]

PACS numbers: 43.28.Ra, 43.28.Dm, 43.28.-g [DKW] Pages: 3489–3496

I. INTRODUCTION

Current efforts to monitor low-level nuclear explosions in the atmosphere and to detect battle-field infrasound have increased interest in separating infrasound from wind noise in outdoor microphone signals. This wind noise increases dramatically with decreasing frequency. For more than 50 years pipe arrays have been employed to average pressure variations over an extended surface area and thus separate infrasound from wind noise.[1–3] Because sound travels approximately 100 times faster than wind, pressure ports in the pipe arrays can be separated far enough to get random signals from the wind and still get coherent signals from sound. If the pressure variations due to the wind are incoherent, the averaging process will reduce the pressure spectrum by a factor of $1/n^{1/2}$ or the power density spectrum by a factor of $1/n$, where n is the number of ports. To design such arrays it is necessary to know how the correlation of the wind-induced pressure variations depends on the sensor separation.

This paper reports wind noise measurements using a three-axis array of 28 low-frequency sensors. By recording the signals from the individual sensors it is possible to study the correlation as a function of sensor separation. Simultaneous measurements were made of the three components of the wind velocity.

In order to minimize the dependence upon wind speed, it is convenient to express the sensor separation in wave numbers. For this purpose the wind velocity is decomposed into a mean component (the convection velocity) and a fluctuating component. The time-varying signal in a stationary anemometer or microphone is assumed to result from a spatially varying field that is frozen in time moving across the anemometer at a velocity equal to the convection velocity (the Taylor hypothesis). The cyclic frequency (F) of the time-varying signals from a stationary sensor is related to the component of the wave number (k_1) in the direction of the convection velocity of the spatially varying field by the equation

$$F = k_1 v/2\pi, \tag{1}$$

where v is the convection velocity.

II. THE MEASURING SYSTEM

A. The infrasound sensors

The infrasound sensors used in this study are of an original design and cover a frequency range of 0.1 to 200 Hz. They are made from piezoelectric bimorphs potted in polyurethane and housed in sections of PVC pipe. Figure 1(c) is a photograph of one of these sensors. Figures 1(a) and (b) picture the internal components. The housing [Fig. 1(c)] is made from a 2.5-in. section of 2-in. O.D. schedule 40 PVC pipe with end caps. Twenty-five 1.9-mm holes are bored in each end cap and four around the center of the cylindrical body. The pressure-sensing elements [Figs. 1(a) and (b)] consist of two pressure sensitive capsules made from four commercially available piezoelectric bimorphs. The bimorphs are 3.5 cm in diameter and have a resonant frequency of about 1.5 kHz. The capsules are constructed by cementing these bimorphs to each side of a brass ring. In order to compensate for temperature changes, one of the capsules has the piezoceramic surface turned to the outside of the capsule and the other has the ceramic turned to the inside. The two capsules are then wired in series. When wired in this way voltage changes produced by the thermal expansion and contraction of the bimorphs cancel and those produced by the bending of the bimorphs in response to pressure changes add. To further thermally insulate the sensors, the capsules are potted in polyurethane, wrapped in fiberglass, and enclosed in a PVC housing. In this configuration, the sensors had a negligible seismic response.

[a)]Portions of this work were presented in a paper entitled "The use of an infrasound microphone array to study wind noise spectra and correlation" at the 145th meeting of the Acoustical Society of America in Nashville, TN, 28 April–2 May 2003.

[b)]Electronic mail: dshields@mil-tec.com

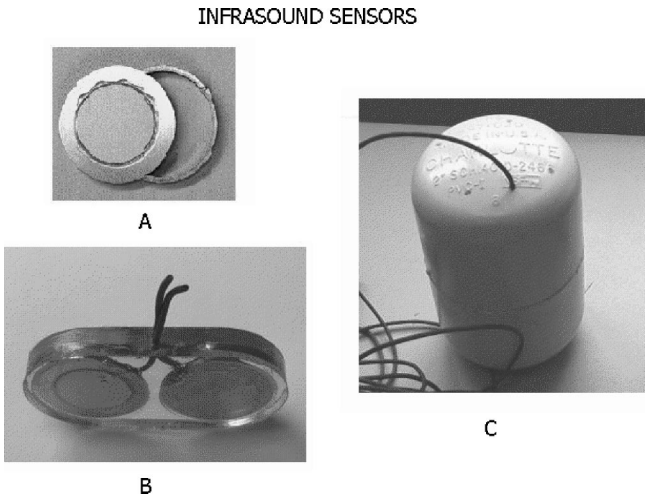


FIG. 1. Infrasound sensors. (a) pictures a piezoelectric ''bimorph'' formed by cementing a piezo-ceramic disc to a brass disc. A pressure sensitive capsule is formed by cementing one of these bimorphs to each side of the brass ring also shown. (b) pictures two capsules, one with the piezo-ceramic turned out and one with it turned in, potted in transparent polyurethane. (c) pictures the housing for the potted capsules. It is made from 2-in. schedule 40 PVC pipe with end caps. The potted capsules are wrapped in fiberglass and enclosed in this PVC housing. Fifty-four holes are bored in the PVC housing as explained in the text.

These sensors are high impedance, charge-generating devices and, as explained below, their sensitivity will depend upon the input impedance of the circuit used to measure their response. For the measurements reported here, the absolute sensitivity was not as important as the variation from sensor to sensor. The absolute sensitivity of the sensors was checked using a number of different techniques. For frequencies from 0.1 to 1 Hz, they were moved up and down in the atmosphere, being careful to avoid effects of acceleration and flow velocity of the air by the sensor. This method checked the response between 0.1 and 1 Hz (vertical displacement of 1 m in the atmosphere under standard conditions produces a change in pressure of 11.8 Pa). For frequencies between 0.1 and 10 Hz, they were placed in an airtight chamber fitted with a moveable piston. The pressure variation was determined by assuming that the volume changes produced by the piston motion were isothermal. For frequencies from 10 up to 100 Hz the response was compared to that of a model 4190 $\frac{1}{2}$-in. B & K microphone in sealed chambers of two different sizes. From these measurements, the absolute sensitivity using the detecting circuit described below was determined to be within 2 dB of 2 mv/Pa for frequencies between 0.1 and 100 Hz.

As for the variation from sensor to sensor, the individual sensors were compared to a reference sensor in a closed chamber. Only two of the sensors varied from the average by more than 12%. Each of the sensors was multiplied by a calibration constant so that all of the sensors had the same sensitivity to within $\pm 3\%$ between 0.1 and 10 Hz. Their phase response over this frequency interval was uniform to ± 0.1 rad.

Twenty-eight of these sensors have been arranged into a three-axis orthogonal array. Counting the one shared sensor, each arm contains ten sensors. The sensors were 0.61 m (2 ft) apart. The individual sensors in each arm were taped to pipes 3.2 cm in diameter. A Campbell Instrument Company CSAT 3 ultrasonic anemometer was mounted 3 m off the ground about 0.5 m from one of the pressure sensors on the vertical arm of the array.

Three analog outputs of the anemometer gave voltages that were proportional to the x, y, and z components of the wind velocity. These voltages were output (updated) 60 times a second and were read with the same data acquisition card as the sensor voltages. By using the analog output of the anemometer the velocity measurements were synchronized with the pressure measurements with a 0.033-s delay.

B. The data acquisition system

The sensor capsules described above are charge-generating devices with a capacitance of approximately 130 nF. As such, they require a high-input impedance detecting circuit. The voltages from the array were measured by connecting the sensors to the input of a National Instrument Company 6031E data acquisition card. This card has 64 single-ended or 32 differential input channels with an input impedance of 100 GΩ in parallel with a 100-pF capacitance. Its maximum sampling rate is 100 000 Hz. In order to avoid cross talk between successively read channels with this card when the sources have high internal impedance, it is necessary to read a shorted input between each two channel readings. This effectively increases the input capacitance of the digitizer by the sampling frequency. The metalized ceramic surfaces of the two capsules in the temperature-compensated transducers were wired to the differential inputs of the data acquisition board with the brass backing plates grounded. A 20-mΩ input shunting resistance was connected from each of the differential inputs to ground to limit the response for frequencies below 0.05 Hz where temperature fluctuations become a problem. The analog channels are scanned 500 times per second. For a single data file, voltage measurements from all of the channels were accumulated in a computer for 120 s. With this digitizer and sampling rate the sensors had a sensitivity of 2.0 mV/Pa between 0.05 and 10 Hz.

III. THE SITES FOR THE MEASUREMENTS

The measurements reported here were made at three different sites. For each of them there were several hundred meters of unobstructed terrain in the direction of the prevailing wind. Figure 2 shows a photograph of the array deployed at site one, a sod field located about 8 miles south of Oxford, MS. The second site was a disked field (the soil surface broken up with clods a few inches in linear dimensions), and the third was a cotton field. The cotton had been picked, but the cotton stalks remained with a uniform height of about 70 cm. The prevailing wind was blowing across the cotton rows. In the data displayed in the figures below, six 2-min runs in the cotton field and the disked field and 15 2-min runs in the sod field have been averaged. The 15 runs in the sod field were taken on two different days under different weather

FIG. 2. The infrasound array as deployed in the sod field which was one of the sites for the measurements.

conditions. The convection velocities as measured 3 m above the ground varied between 4 and 8 m/s, with most of the measurements falling between 5 and 7 m/s.

IV. THE AVERAGE RESPONSE FROM EACH ARM OF THE ARRAY

The power spectral density of the pressure variations for each of the three arms of the array for each of the three sites is plotted in Fig. 3. For this purpose, the 2-min data files for each run were Fourier analyzed using the MATLAB ''Spectrum'' program. The fast Fourier transform (FFT) calculation used 10 000 points (at the sampling rate of 500 Hz this corresponds to 20 s), an overlap of 5000 points, and a Hanning window. The spectra have not been normalized by dividing by 1/2 the sampling rate. Therefore, the area under the curves represents the variance of the pressure signals. The data for these curves were obtained from ten sensors in each arm of the array. The top curve, A, in each figure is obtained by calculating the spectra of time signals from each of the ten sensors in the arm and then averaging these spectra for all the runs at a given site. Therefore, the A curves represent the average spectrum from a single sensor in each arm. The average is for 60 spectrum curves for the disked and cotton fields and 150 for the sod field. For the lower curves (B), the time signals from the nine sensors of each arm of the array were instantaneously summed and divided by the number of sensors to obtain an average time signal. The spectra of these average time signals were calculated for each run and then these 6 or 15 spectra averaged for each site. For these B curves then, if the sensors are far enough apart so that their signals are uncorrelated, this averaging of the time signals should divide the noise power density by the number of sensors. This is seen to be the case for the vertical arm signals at all three sites for frequencies above approximately 1 Hz. For the two horizontal arms that are on the ground the wind noise at the higher wave numbers is reduced to where the spectra are influenced by background sound. Since the sound is at

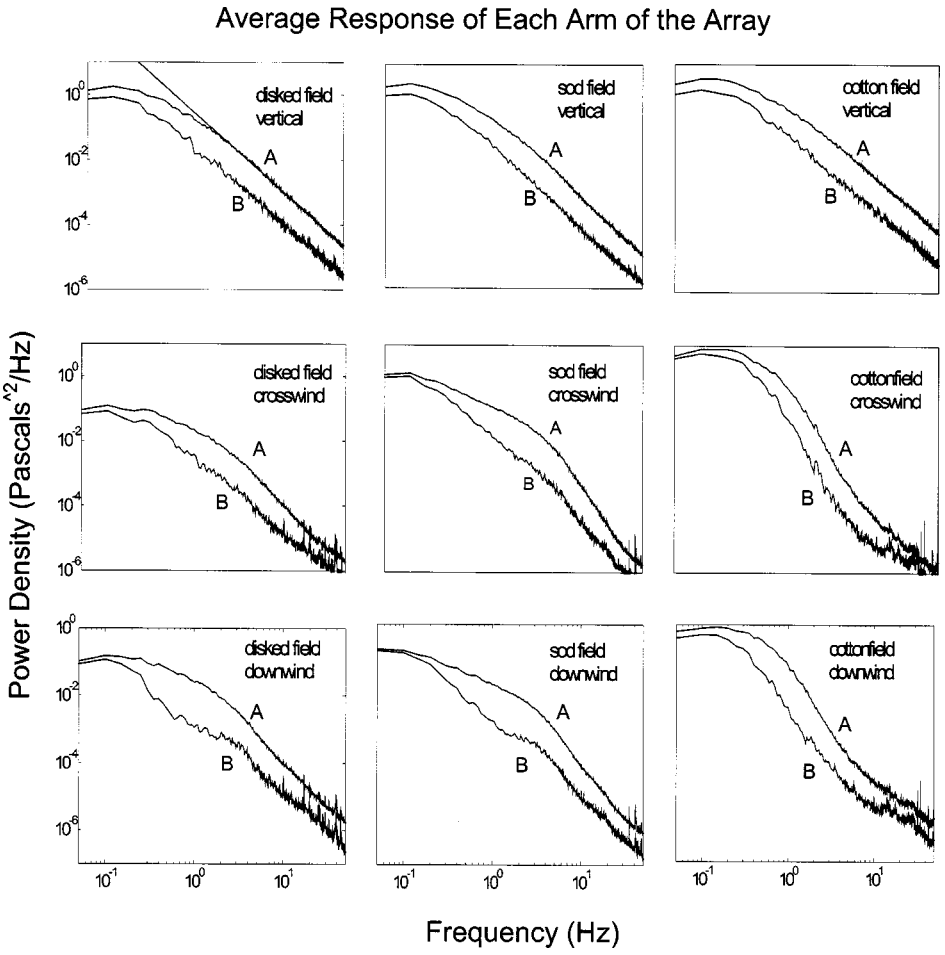


FIG. 3. The power spectral density of the varying pressure as measured by each arm of the array at each of the three sites. Curves A were obtained by averaging the spectra from ten sensors in each arm. Curves B were obtained by taking the spectra of the average signal from the same ten sensors.

least partially correlated over the arm of the array, averaging the time signals reduces the power level by less than the $1/n$ factor.

The average spectrum for the ten sensors in the vertical arm that are from 0 to 18 ft off the ground has very nearly a $-\frac{7}{3}$ slope. This is illustrated by the straight line plotted through the data for the vertical arm in the disked field. For the horizontal arms on the ground, the surface boundary reduces the turbulence and causes a deviation of the spectral slopes from $-\frac{7}{3}$. The spectra for the cotton field data differ significantly from the other two sites. The noise level in the vertical arm is greater and the shape of the spectra for the two arms on the ground is different. The pressure levels at frequencies in the neighborhood of 0.3 Hz are increased while those around 3 Hz are decreased. Other unique aspects of the cotton field data are discussed below.

For the disked field and sod field, averaging the signals from the downwind arm produces greater than $1/n$ noise reduction for wave frequencies between about 0.5 and 5 Hz. This can be explained by assuming that the turbulent pressure field has a periodicity. When the convection velocity carries this periodic field across the downwind arm of the array, for wavelengths comparable to the length of the array arm, signals from different parts of the array will be out of phase and cancel out in the averaging process. This effect is examined more carefully in the next section.

V. CORRELATION STUDIES

It is interesting to compare the measurements made here with those made by J. T. Priestley nearly 40 years ago. He made narrow-band pressure correlation measurements in the frequency range 0.008 to 1 Hz for wind speeds ranging from 2.1 to 7.2 m/s. He used six pressure sensors constructed at the Bureau of Standards to study the dependence of the correlation on the downwind and crosswind separation of the sensors.[4] He determined, in agreement with earlier studies, that the narrow-band correlation coefficients in the down wind and crosswind directions are given by

$$R(\text{downwind}) = e^{-\alpha x}\cos(k_1 x) \tag{2}$$

and

$$R(\text{crosswind}) = e^{-\beta y}, \tag{3}$$

with x and y the downwind and crosswind separations. Priestley found that, though the convection velocity given by $(2\pi^* F/k_1)$ was highly dependent upon meteorological variables, the functional relationships between α, β, and k_1 were remarkably independent of them. (F was the center frequency of his filtered signal and k_1 chosen to fit his data.) He obtained[5]

$$\alpha = 0.33 \cdot k_1^{1.28} \quad \text{for } 0.5 < 1/k_1 < 50, \tag{4}$$

$$1/\beta = 0.84 \cdot (1/\alpha)^{0.74} \quad \text{for } 3 < (1/\alpha) < 500, \tag{5}$$

where α and β, are measured in reciprocal meters and k_1 in radians per meter.

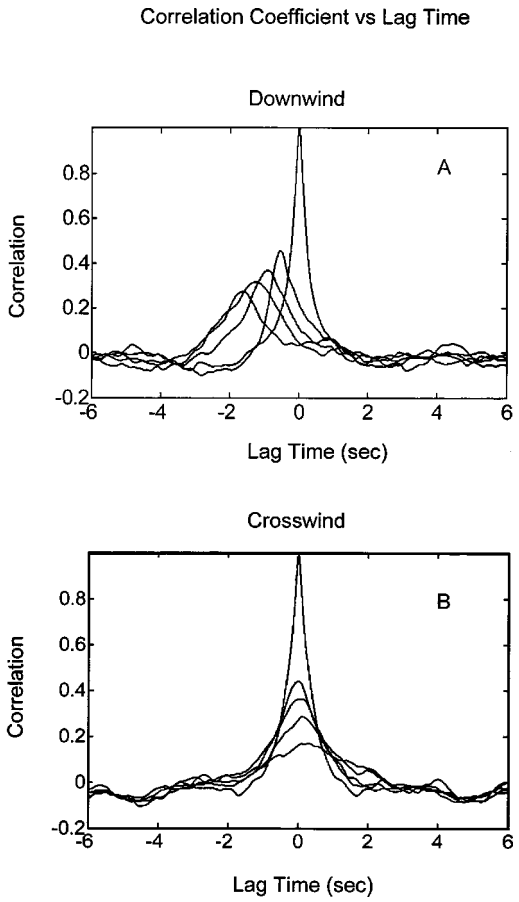


FIG. 4. The correlation coefficient versus the lag time for signals from five sensors in (a) the downwind arm of the array and (b) the crosswind arm of the array. For the crosswind arm the sensors are separated by 2 ft (61 cm) and for the downwind arm they are separated by 4 ft (122 cm).

It should be noted in comparing these constants to those reported here for measurements at larger wave numbers, a value of α proportional to k_1 to the first power fits his data for wave numbers between 0.5 and 2 m^{-1}.

To study the correlation in the data taken here, signals from each arm of the array have been analyzed using the MATLAB XCOV program. This program calculates the cross correlation between the signals as a function of lag time. Figures 4(a) and (b) shows this correlation for signals from five sensors in the downwind and crosswind arms of the array, respectively, for the disked field data. Plotting the correlation at zero lag time versus sensor separation in these two figures gives the decrease in correlation with distance in the crosswind and downwind directions. Plotting the peak heights versus lag time for the downwind arm [Fig. 4(a)] gives the decrease in correlation with time at a fixed position in the turbulent pressure pattern (according to the Taylor hypothesis). As has been pointed out by Bass, Raspet, and Messer, the convection velocity should equal to the slope of the plot of the sensor separation versus lag time at which the peak of the curves in Fig. 4(a) occurs.[6]

It is interesting to look at these correlation curves as a function of frequency. For this purpose the digitized time domain signals are filtered with a fourth-order bandpass Butterworth filter with the high- and low-pass frequencies the same. Figures 5(a)–(c) show the correlation of the signals at

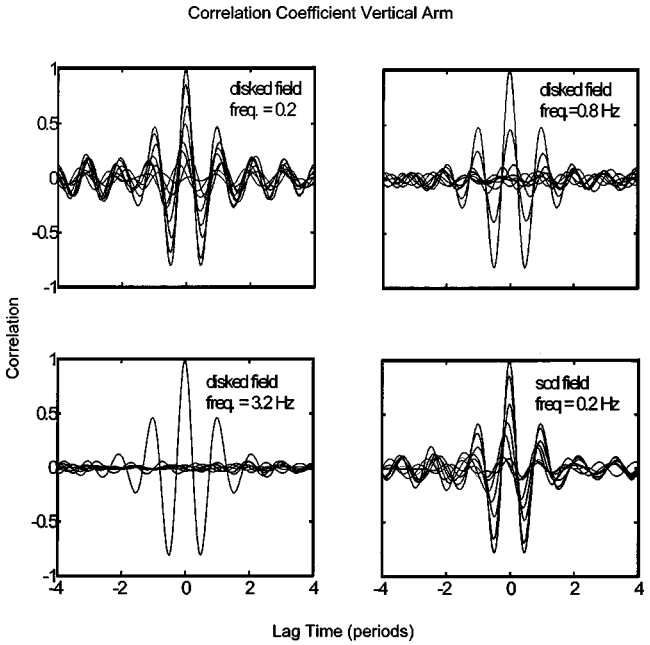


FIG. 5. Single-frequency correlation versus lag time in periods for varying pressure signals from eight sensors in the vertical arm of the array.

three different frequencies from eight sensors in the vertical arm of the array for the disked field data. The sinc function shape of the correlation curves is characteristic of the correlation of a narrow band of frequencies. The lag time has been normalized by dividing by the period of the filtered signal. This gives the same shape to curves at different frequencies. As the frequency is increased from 0.2 Hz in Fig. 5(a) to 3.2 Hz in Fig. 5(c), the rate of decrease in correlation with distance rapidly increases to where at 3.2 Hz there is negligible correlation even between the adjacent sensors that are separated by only 0.6 m. Figure 5(d) gives the curves for the same frequency as Fig. 5(a) but for the sod field. The correlation for the two sites is very similar.

Figure 6 is a plot of the correlation at zero lag time between sensors in the downwind arm of the array versus the sensor separation in wavelengths. Data for the disked, sod, and cotton fields for a number of frequencies have been plotted on the same graph. The data for each site have been

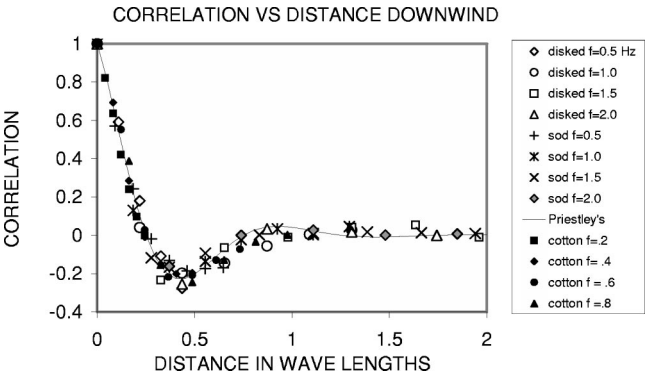


FIG. 6. Correlation of the varying pressure signals versus the sensor separation in wavelengths in the downwind direction. The data are for four different frequencies and for three different sites as designated in the legend. The solid curve is the fit to Priestley's data taken at much lower frequencies in 1964.

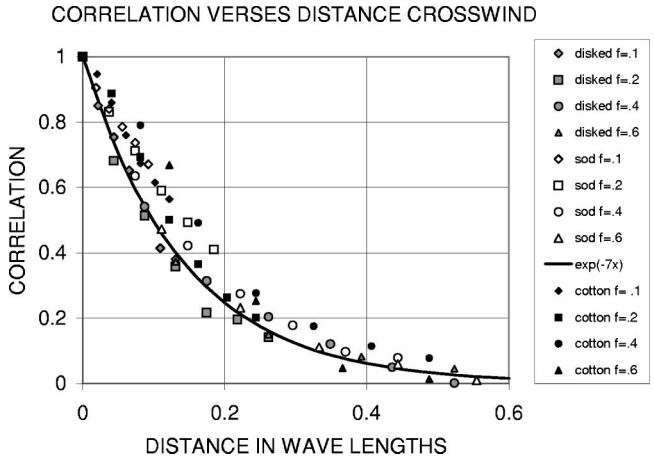


FIG. 7. Correlation of the varying pressure signals versus the sensor separation in wavelengths in the crosswind direction. The data are for four different frequencies and for three different sites as designated in the legend. The solid exponential curve is drawn visually to fit the data.

averaged not only for all runs at the site, but also for all pairs of sensors with the same separation. The wavelength was determined by dividing the convection velocity by the frequency of the filter. However, since the sensors are on the ground, the anemometer measurement of the velocity made 3 m above the ground is inadequate because of near-surface wind shear. Instead the convection velocity for the three sites was adjusted to make the data fit Eq. (2). These velocities were 2.8, 3.3, and 3.0 m/s for the disked, sod, and cotton fields, respectively. The velocities obtained in this way were approximately equal to those obtained from the plot of the sensor separation versus the lag times at which the maximum in the correlation occurs (see Fig. 4). The average velocities for the three sites measured with the anemometer 3 m above the ground were 6.2, 6.7, and 5.9 m/s, respectively.

To within experimental error, the correlation is seen to be a function of the separation divided by the wavelength independent of frequency, wind velocity (over the limited range of velocities measured), and even the terrain. Priestley references a number of authors who observed this functional relationship back in the 1960s.[4] The customary plot is correlation verses 2π times the abscissa in Fig. 6.

It is interesting to compare Priestley's measurements made many years ago in a grassy field near Dulles International Airport, Washington, DC, at much lower frequencies than those reported here. His α as given by Eq. (2) is slightly frequency dependent. The α value used to draw the curve in Fig. 6 corresponds to Priestley's α as calculated from Eq. (4) with a wave number of 4.7 m^{-1}. This is the highest wave number for the data plotted in Fig. 6.

Also of interest is the variation in the correlation coefficient with sensor separation in the crosswind and vertical directions. Figures 7 and 8 plot this dependence for the same runs and same convection velocities as Fig. 6. For Fig. 8 the separation is in the vertical direction between sensors that are 1.2 to 5.5 m above the ground. To within the accuracy of the measurements, the decay can be plotted as a function of the separation in wavelengths independent of frequency for the disked field, but for the sod and cotton fields the low frequencies decay slower in the crosswind direction and faster

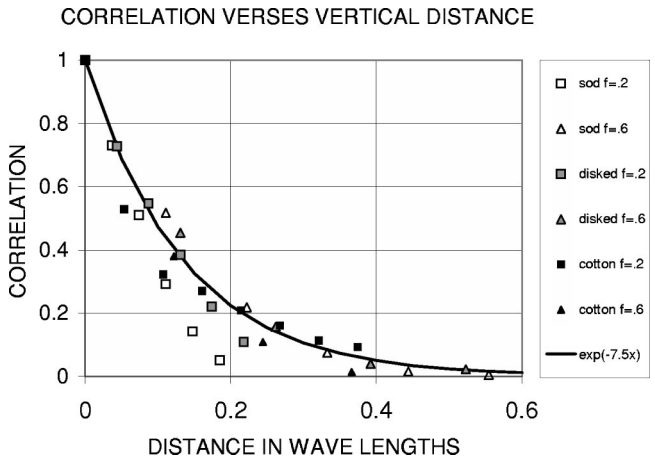


FIG. 8. Correlation of the varying pressure signals versus the sensor separation in wavelengths in the vertical direction. The data are for two different frequencies and for three different sites as designated in the legend.

in the vertical direction. An exponential curve has been drawn visually through the data in Figs. 7 and 8 with a decay constant of approximately -7 per wavelength for the crosswind and -7.5 for the vertical direction. Priestley measured the crosswind decay and got a decay constant that was slightly frequency dependent. For the frequency range reported his value would be about 3.5, or about half of our measurement.

VI. VELOCITY AND PRESSURE SPECTRA

A. Validation of the pressure measurements

Widely different results are reported for pressure spectrum measurements in turbulence.[7–13] One must always be concerned with what is being measured by a pressure sensor in turbulent flow.

Of first concern is the disturbance introduced by the sensing element. Even when this is minimized, Morgan and Raspet conclude, ''The dominant source of wind noise in outdoor microphones is the pressure fluctuations caused by the velocity fluctuations of the incoming flow.'' [13] To study this question further, sensors employed here were oriented in different directions relative to the wind direction and the spectra of their response compared. This response was found be insensitive to orientation relative to wind direction over the frequency range of interest. In addition, their response in wind has been compared to the response obtained using a ''Quad Disk.'' [14] This device is designed to measure pressure independent of wind direction and magnitude. Bedard *et al.* have used it to measure pressure variations in a 0.5- to 2.5-Hz band over a 3-month period.[15] Their results differ significantly from one period of measurement to another. They plot rms pressure variations in Pascals versus wind speed in meters per second. Their data for two different sample periods are fit with two equations, $dp = 0.0036\ \mathrm{U}^{3.02}$ and $dp = 0.0063\ \mathrm{U}^{1.98}$, where the pressure is in Pascals and the velocity in meters per second. [Bedard's Eq. (15) for his Dec. 11 data evidently has a factor of 2 missing.] Figure 9 shows a similar plot of measurements made here with the single sensor close to the anemometer in the vertical arm of

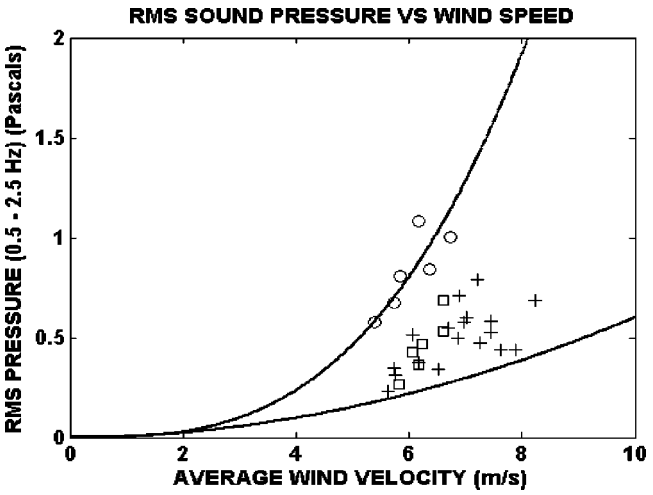


FIG. 9. The root mean square pressure variations in the frequency interval between 0.5 and 2.5 Hz plotted versus the average wind speed. Data are shown for six runs in the cotton field, six runs in the disked field, and 18 runs in the sod field. Circles are for the cotton field, squares are for the disked field, and crosses are for the sod field.

the array averaged over the same frequency interval. The solid curves are plots of the two functions that fit Bedard's two sets of data.

As a further comparison of the response of our sensors with that of the Quad Disk, a B&K $\frac{1}{2}$-in. microphone was sealed in the axial tube of the Quad Disk and the Quad Disk and one of our sensors placed at a height of 3 m in the wind at the local airport. The power density spectra of the two devices are shown to agree remarkably well in Fig. 10. In this figure, 25 min of data were analyzed as explained in the discussion of Fig. 3. The average wind velocity was 4.6 m/s.

B. The wavenumber dependence of the pressure and velocity spectra

Spectra of turbulence are generally explained by assuming that energy is fed into the turbulence at small wave numbers and is dissipated by viscous losses at high wave num-

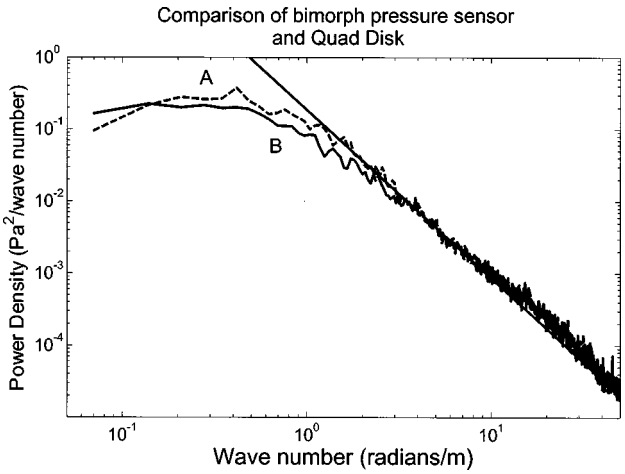


FIG. 10. The comparison of the power spectral density of 25 min of wind noise data taken with (a) the quad disk (dashed line) and (b) one of the sensors used to take the data reported here (solid line). Both were approximately 3 m off the ground and approximately 5 m apart. The straight line through the data has a slope of $-\frac{7}{3}$. The wind speed was 4.7 m/s.

bers. The energy cascades from large eddies to smaller ones. For a homogeneous turbulent field there will be a range of wave numbers called the inertial subrange where there is negligible production and dissipation of turbulent energy. In this range of wave numbers the one-dimensional velocity spectrum from symmetry considerations is expected to obey the equation[16]

$$F_1(k_1) = a_v k_1^{-5/3}. \tag{6}$$

Here F_1 is the power spectral density of the component of the varying velocity in the direction of the convection velocity, and a_v is independent of frequency but depends upon the rate of energy dissipation in the turbulent field. If the turbulence is isotropic over this range of wave numbers, the spectra for the other two velocity components in the k_1 direction have a similar form with a_v multiplied by $\frac{4}{3}$ and have a negligible correlation with each other. The three conditions, $-\frac{5}{3}$ power law, $\frac{4}{3}$ ratio between the transverse and longitudinal velocity components, and the vanishing (or very low) co spectral levels, are used as a test for the existence of an inertial subrange.[16]

From dimensional analysis the power density of the dynamic pressure (i.e., p/ρ) in the inertial subrange is expected to vary as[17]

$$F(P/\rho) = a_p k_1^{-7/3}, \tag{7}$$

with a_p independent of frequency but dependent upon the rate of energy dissipation in the turbulent field. However, the exponent of the wave number is contested by both experiment and theory.[7,11]

The range of wave numbers over which the turbulent fields studied here might be considered homogeneous, and therefore the range over which the inertial subrange equations might be expected to apply, is limited on the high side by the dimension of the sensor and anemometer (approximately 0.1 m) and on the low side by the height the sensor and anemometer were above the ground (approximately 3 m). The spectra for the sensors on the ground would not be expected to demonstrate the inertial subrange frequency dependence. These conditions would limit the inertial subrange to wave numbers between approximately $2\pi/0.2$ and $2\pi/6$ m, or between 30 and 1 rad/m.

Figure 11 shows the averaged power spectral densities calculated using MATLAB software as explained in discussing Fig. 3 above. The A curves are for the variations in the kinematic pressure (i.e., pressure divided by the density of air). The B curves are for the variations in the component of the velocity in the direction of the convection velocity. The pressure variations were measured by the single sensor located about 0.5 m from the anemometer. Both were approximately 3 m off the ground. The spectra for each 2-min run were calculated in units of per radian of distance by specifying the sampling rate in samples per radian of distance which is equal to 2π times the samples per second divided by the convection (average) velocity for each run. These spectra (6 for the disked and cotton fields and 18 for the sod field) were averaged and plotted in Fig. 11. This plot of the component of the varying velocity in the direction of the convection velocity and the pressure divided by the air density against

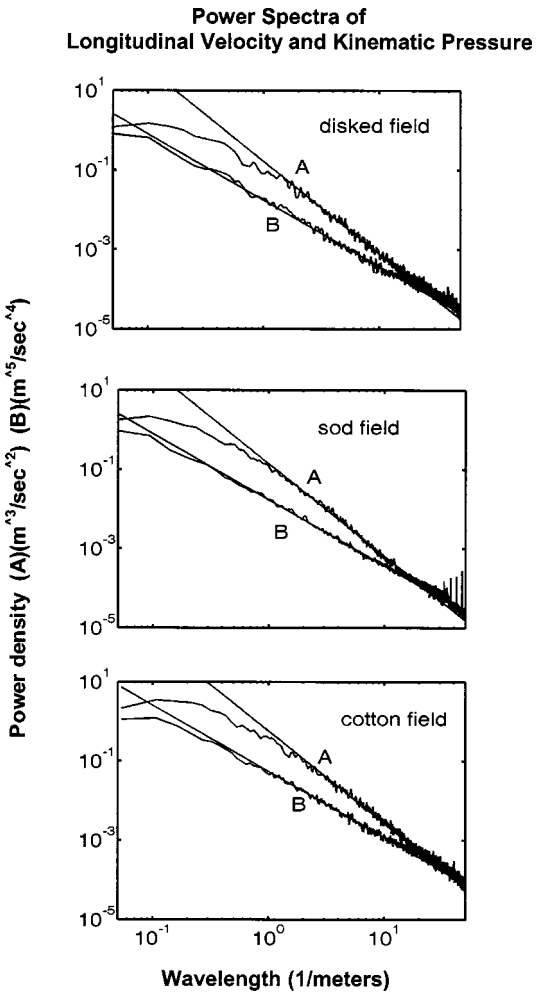


FIG. 11. The power density spectra for the varying wind velocity and pressure as measured at three different sites. Curves A are for velocity and curves B for pressure.

the wave number in the direction of the convection velocity was used because these are the quantities frequently calculated theoretically. The curves would be changed insignificantly if the spectrum of the total varying velocity were plotted.

The straight lines drawn through the curves have slopes of $-\frac{5}{3}$ for the velocities and $-\frac{7}{3}$ for the pressures. The constant multiplying factors are adjusted to give the best fit between 1 and 10 m^{-1}.

As a further test for the inertial subrange, similar curves were plotted for the transverse components of velocity. As stated above, if the turbulent field is isotropic, the ratio of the power density of the transverse to the longitudinal components should equal $\frac{4}{3}$. For the average of 36 min of sod field data and 12 min of the disked field data, this ratio is approximately equal to $\frac{4}{3}$. However, for the average of 12 min of cotton field data this ratio is approximately 1. These are the same data used in the correlation studies. For all three sites the ratio of the power density of the vertical component to the downwind component is approximately 1. The third condition for the inertial subrange, i.e., small or vanishing cospectra between the different velocity components, is met by the x and y velocity components in the sod and disked fields, but the x and y components in the cotton field and the x and

z components at all three cites showed a maximum correlation of approximately 0.2.

VII. CONCLUSIONS

(1) The noise power reduction produced by averaging signals from nine individual sensors varied as 1/(number of sensors) for a range of frequencies when the sensor spacing was greater than a few tenths of the wavelength in the turbulent pressure field, where the wavelength is determined by dividing the convection wind velocity by the frequency.

(2) The frequency-dependent correlation of the wind noise over a range of wind velocities and atmospheric and environmental conditions varies as $\exp(-3.2x)\cos(2\pi x)$, where x is the sensor separation in wavelengths in the down wind direction. This allows greater than $1/n$ noise reduction if the aperture of the array is greater than $\frac{1}{4}$ wavelength of the pressure variations in the wind.

(3) The correlation as a function of sensor separation for arrays in the crosswind and vertical directions is given approximately by $\exp(-7x)$ for a range of wind velocities and atmospheric and environmental conditions.

(4) For frequencies expected to fall within the inertial subrange, the $-\frac{5}{3}$ law for the velocity spectrum and the $-\frac{7}{3}$ law for the pressure spectrum were found to hold.

ACKNOWLEDGMENTS

The help of Chris Clark and Brandon Smith in taking data is gratefully acknowledged. Thanks also go to Dr. Ken Gilbert, Dr. Carrick Talmadge, Dr. Richard Raspet, Dr. Garth Frazier, and Dr. Henry Bass for helpful discussions, and to the U.S. Army Armament Research Development and Engineering Center for supporting this work.

[1] F. B. Daniels, "Noise-reducing line microphone for frequencies below 1 cps," J. Acoust. Soc. Am. **31**, 529–531 (1959).

[2] S. M. Tenney, B. T. Mays, and J. M. Noble, "A methodology for infrasonic detection and array signal processing," Army Research Laboratory, Security Office, 2800 Powder Mill RD. Adelphi, MD 20783 (1999).

[3] M. A. Hedlin, B. Alcoverro, and G. D'Spain, "Evaluation of rosette infrasonic noise reducing spatial filters," J. Acoust. Soc. Am. **114**, 1807–1820 (2003).

[4] J. T. Priestley, "Correlation Studies of Pressure Fluctuations on the Ground Beneath a Turbulent Boundary Layer," National Bureau of Standards Report No. 8942 (1965); also M.S. thesis, University of Maryland (1965).

[5] J. T. Priestley, "Calculation of the Effectiveness of Infrasonic Line Microphones for Reducing Wind Noise," National Bureau of Standards Report 9380 (1966). This report corrects a mistake in the value of alpha given in the 1965 report.

[6] H. E. Bass, R. Raspet, and J. O. Messer, "Experimental determination of wind speed and direction using a three microphone array," J. Acoust. Soc. Am. **97**, 695–696 (1995).

[7] W. K. George, P. D. Beuther, and R. E. A. Arndt, "Pressure spectra in turbulent free shear flows," J. Fluid Mech. **148**, 155–191 (1948).

[8] J. M. Wilczak, S. P. Oncley, and A. J. Bedard, Jr., "Turbulent pressure fluctuations in the atmospheric surface layer," *Proc. 10th Symposium on turbulence and diffusion*, Portland, OR (Am. Meteor. Soc., Boston, MA, 1992, pp. 167–170).

[9] J. D. Albertson, G. G. Katul, M. B. Parlange, and W. E. Eichinger, "Spectral scaling of static pressure fluctuations in the atmospheric surface layer: the interaction between large and small scales," Phys. Fluids **10**, 1725–1732 (1998).

[10] Y. Tsuji and T. Ishihara, "Similarity scaling of pressure fluctuation in turbulence," Phys. Rev. E **68**, 026309 (2003).

[11] A. I. Grachev and M. I. Mordukhovich, "Spectrum of pressure fluctuations in the atmospheric ground layer," Izv., Acad. Sci., USSR, Atmos. Oceanic Phys. **24**, 159–160 (1988).

[12] H. V. Fuchs, "Measurement of pressure fluctuations within subsonic turbulent jets," J. Sound Vib. **22**, 361–378 (1972).

[13] S. Morgan and R. Raspet, "Investigation of the mechanisms of low frequency wind noise generated outdoors," J. Acoust. Soc. Am. **92**, 1180–1183 (1992).

[14] R. T. Nishiyama and A. J. Bedard Jr., "Quaddisc static pressure probe for measurements in adverse atmospheres: With a comparative review of static pressure probe designs," Rev. Sci. Instrum. **62**, 2193–2204 (1991).

[15] A. J. Bedard, R. W. Whitaker, G. E. Greene, P. Mutschlecner, R. T. Nishiyama, and M. Davidson, "Measurements of pressure fluctuations near the surface of the earth," *Proc. 10th Symposium on turbulence and diffusion*, Portland, OR (Am. Meteor. Soc., Boston, MA, 1992, pp. 293–296).

[16] J. C. Kaimal and J. J. Finnigan, *Atmospheric Boundary Layer Flows* (Oxford U. P., Oxford, 1994), pp. 36 and 37.

[17] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT, Cambridge, 1975), Vol. 2, p. 374.

Modeling Rayleigh and Stoneley waves and other interface and boundary effects with the parabolic equation

Wayne Jerzak[a] and William L. Siegmann
Rensselaer Polytechnic Institute, Troy, New York 12180

Michael D. Collins[b]
Naval Research Laboratory, Washington, DC 20375

(Received 27 September 2004; revised 17 February 2005; accepted 18 February 2005)

An improved approach for handling boundaries, interfaces, and continuous depth dependence with the elastic parabolic equation is derived and benchmarked. The approach is applied to model the propagation of Rayleigh and Stoneley waves. Depending on the choice of dependent variables, the operator in the elastic wave equation may not factor or the treatment of interfaces may be difficult. These problems are resolved by using a formulation in terms of the vertical displacement and the range derivative of the horizontal displacement. These quantities are continuous across horizontal interfaces, which permits the use of Galerkin's method to discretize in depth. This implementation extends the capability of the elastic parabolic equation to handle arbitrary depth dependence and should lead to improvements for range-dependent problems. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1893245]

PACS numbers: 43.30.Bp, 43.30.Ma [DRD] Pages: 3497–3503

I. INTRODUCTION

The parabolic equation method is a powerful approach for solving range-dependent wave propagation problems. A parabolic wave equation is derived by factoring the operator in an elliptic wave equation into a product of incoming and outgoing operators and assuming that outgoing energy dominates. One-way solutions are exact in stratified media and accurate if the medium varies gradually in the horizontal. Efficiency gains of several orders of magnitude can be achieved by approximating an elliptic wave equation with a parabolic wave equation. The parabolic equation method has important applications in ocean acoustics.[1] In many cases, there is interaction with the ocean bottom, which may support shear waves.

The development of the elastic parabolic equation has proven to be more difficult than the development of the acoustic parabolic equation. After the parabolic equation method was first applied to ocean acoustics,[2–4] it was more than a decade before formulation, stability, and accuracy issues were resolved and the elastic parabolic equation was successfully implemented.[5–8] There remain unresolved issues in this area, including improving accuracy for range-dependent problems. Since the factorization of the operator is approximate for range-dependent problems, corrections based on energy conservation,[9–12] single scattering,[13–15] and other ideas[16] have been developed. These approaches provide improved accuracy, but the problem has not been fully resolved. A key component of the problem is to handle depth dependence effectively. There are several choices for the dependent variables. For some choices, the operator does not factor, which makes the formulation unsuitable for the elastic parabolic equation. For other choices, the variables are not continuous across interfaces, which causes difficulties in the depth discretization and for range-dependent problems.

In this paper, we extend the elastic parabolic equation to accurately and efficiently handle an arbitrary depth dependence, including boundaries, interfaces, and continuous variations. The dependent variables are the vertical displacement and the horizontal derivative of the horizontal displacement, which have been used to derive an elastic parabolic equation for anisotropic media.[17] It is possible to factor the operator in these variables, which are continuous across horizontal interfaces. In this formulation, problems involving continuous depth dependence, solid-solid interfaces, and free boundaries can be discretized efficiently using Galerkin's method.[18] Fluid-solid interfaces can be handled by explicitly enforcing the interface conditions. We describe the approach, demonstrate its accuracy, and apply it to problems involving Rayleigh and Stoneley waves.

II. THE PARABOLIC EQUATION

In this section, we derive the elastic parabolic equation. Following the standard approach for parabolic equation derivations, we treat the medium as range independent (stratified) and work in Cartesian coordinates. In practice, range dependence is treated by making corrections across vertical interfaces and cylindrical geometry is treated by factoring out a spreading term. We consider a two-dimensional problem, where z is the depth and the range x is the horizontal distance from a source. The compressional and shear wave speeds c_p and c_s are related to the density ρ and the Lamé parameters λ and μ by $\rho c_p^2 = \lambda + 2\mu$ and $\rho c_s^2 = \mu$. We work in the frequency domain and remove the factor $\exp(-i\omega t)$ from the dependent variables, where ω is the circular frequency and t is time. Attenuation is taken into account by

[a]Present Address: Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02420.
[b]Author to whom correspondence should be addressed. Electronic mail: collins@noddy.nrl.navy.mil

replacing c_p and c_s with $c_p(1+i\eta\beta_p)^{-1}$ and $c_s(1+i\eta\beta_s)^{-1}$, where β_p and β_s are the compressional and shear attenuations in decibels per wavelength and $\eta^{-1}=40\pi\log_{10}e$.

In an isotropic medium, the displacements u and w and stresses σ_{xx}, σ_{xz}, and σ_{zz} satisfy the momentum equations,[19]

$$\frac{\partial\sigma_{xx}}{\partial x}+\frac{\partial\sigma_{xz}}{\partial z}+\rho\omega^2 u=0, \tag{1}$$

$$\frac{\partial\sigma_{xz}}{\partial x}+\frac{\partial\sigma_{zz}}{\partial z}+\rho\omega^2 w=0, \tag{2}$$

and the constitutive equations,

$$\sigma_{xx}=(\lambda+2\mu)\frac{\partial u}{\partial x}+\lambda\frac{\partial w}{\partial z}, \tag{3}$$

$$\sigma_{zz}=\lambda\frac{\partial u}{\partial x}+(\lambda+2\mu)\frac{\partial w}{\partial z}, \tag{4}$$

$$\sigma_{xz}=\mu\frac{\partial u}{\partial z}+\mu\frac{\partial w}{\partial x}. \tag{5}$$

Substituting Eqs. (3)–(5) into Eqs. (1) and (2), we obtain

$$(\lambda+2\mu)\frac{\partial^2 u}{\partial x^2}+\frac{\partial}{\partial z}\left(\mu\frac{\partial u}{\partial z}\right)+\rho\omega^2 u+(\lambda+\mu)\frac{\partial^2 w}{\partial x\partial z}+\frac{\partial\mu}{\partial z}\frac{\partial w}{\partial x}=0, \tag{6}$$

$$\mu\frac{\partial^2 w}{\partial x^2}+\frac{\partial}{\partial z}\left((\lambda+2\mu)\frac{\partial w}{\partial z}\right)+\rho\omega^2 w+(\lambda+\mu)\frac{\partial^2 u}{\partial x\partial z}+\frac{\partial\lambda}{\partial z}\frac{\partial u}{\partial x}=0. \tag{7}$$

These equations are in the form,

$$\left(\tilde{L}\frac{\partial^2}{\partial x^2}+\tilde{M}\right)\begin{pmatrix}u\\w\end{pmatrix}=\tilde{N}\frac{\partial}{\partial x}\begin{pmatrix}u\\w\end{pmatrix}, \tag{8}$$

where $\tilde{L}$, $\tilde{M}$, and $\tilde{N}$ are depth operators.

Due to the presence of the term on the right-hand side of Eq. (8), the operator does not readily factor into a product of incoming and outgoing operators. The equations of elasticity are in the form of Eq. (8) for any choice of dependent variables, but the term on the right-hand side vanishes for some choices, such as the (Δ,w) formulation,[5] where

$$\Delta\equiv\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z} \tag{9}$$

is the dilatation. Since Δ is not continuous across interfaces, the (Δ,w) formulation is not suitable for an implementation based on Galerkin's method. In this paper, we consider the (u_x,w) formulation, where

$$u_x\equiv\frac{\partial u}{\partial x}. \tag{10}$$

Taking the range derivative of Eq. (6) and substituting Eq. (10) into Eqs. (6) and (7), we obtain

$$(\lambda+2\mu)\frac{\partial^2 u_x}{\partial x^2}+\frac{\partial}{\partial z}\left(\mu\frac{\partial u_x}{\partial z}\right)+\rho\omega^2 u_x+(\lambda+\mu)\frac{\partial^3 w}{\partial x^2\partial z}+\frac{\partial\mu}{\partial z}\frac{\partial^2 w}{\partial x^2}=0, \tag{11}$$

$$\mu\frac{\partial^2 w}{\partial x^2}+\frac{\partial}{\partial z}\left((\lambda+2\mu)\frac{\partial w}{\partial z}\right)+\rho\omega^2 w+(\lambda+\mu)\frac{\partial u_x}{\partial z}+\frac{\partial\lambda}{\partial z}u_x=0. \tag{12}$$

These equations are in the form,

$$\left(L\frac{\partial^2}{\partial x^2}+M\right)\begin{pmatrix}u_x\\w\end{pmatrix}=0, \tag{13}$$

where L and M are depth operators. Multiplying Eq. (13) by L^{-1}, we obtain

$$\left(\frac{\partial^2}{\partial x^2}+L^{-1}M\right)\begin{pmatrix}u_x\\w\end{pmatrix}=0. \tag{14}$$

Factoring the operator in Eq. (14), we obtain

$$\left(\frac{\partial}{\partial x}+i(L^{-1}M)^{1/2}\right)\left(\frac{\partial}{\partial x}-i(L^{-1}M)^{1/2}\right)\begin{pmatrix}u_x\\w\end{pmatrix}=0. \tag{15}$$

Assuming that outgoing energy dominates incoming energy, we obtain the (u_x,w) parabolic equation,

$$\frac{\partial}{\partial x}\begin{pmatrix}u_x\\w\end{pmatrix}=i(L^{-1}M)^{1/2}\begin{pmatrix}u_x\\w\end{pmatrix}. \tag{16}$$

Since the (u_x,w) parabolic equation is identical to the (Δ,w) parabolic equation of Ref. 20 (with different definitions for L and M), it can be implemented using the split-step Padé solution, which is based on the formal solution,

$$\begin{pmatrix}u_x(x+\Delta x)\\w(x+\Delta x)\end{pmatrix}=\exp(i\Delta x(L^{-1}M)^{1/2})\begin{pmatrix}u_x(x)\\w(x)\end{pmatrix}, \tag{17}$$

where Δx is the range step. For a source in a fluid layer, the self-starter as described in Refs. 21 and 22 can be used to generate an initial condition for Eq. (17). The case of a source in a solid layer is described in Sec. III. The exponential of the operator in Eq. (17) can be approximated using a rational approximation as described in Ref. 20 and discretized in depth using Galerkin's method as described in Sec. IV.

III. THE SELF-STARTER

In this section, we generalize the self-starter to the case of a source in an elastic layer in the (u_x,w) formulation. We first consider a homogeneous medium and apply the representation,

$$u=\frac{\partial\phi}{\partial x}+\frac{\partial\psi}{\partial z}, \tag{18}$$

$$w=\frac{\partial\phi}{\partial z}-\frac{\partial\psi}{\partial x}, \tag{19}$$

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial z^2}+k_p^2\phi=0, \tag{20}$$

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial z^2}+k_s^2\psi=0, \tag{21}$$

where $k_p=\omega/c_p$, $k_s=\omega/c_s$, and ϕ and ψ are the compressional and shear potentials.

Placing a term corresponding to a compressional source P at $z=z_0$ on the right-hand side of Eq. (20) and differentiating Eq. (21), we obtain

$$\frac{\partial^2}{\partial x^2}\begin{pmatrix}\phi\\ \psi_x\end{pmatrix}_P+\frac{\partial^2}{\partial z^2}\begin{pmatrix}\phi\\ \psi_x\end{pmatrix}_P+\begin{pmatrix}k_p^2 & 0\\ 0 & k_s^2\end{pmatrix}\begin{pmatrix}\phi\\ \psi_x\end{pmatrix}_P=\begin{pmatrix}-2i\,\delta(x)\,\delta(z-z_0)\\ 0\end{pmatrix}, \tag{22}$$

$$\psi_x\equiv\frac{\partial\psi}{\partial x}. \tag{23}$$

Integrating Eq. (22) about an arbitrarily small interval about $x=0$ and using the symmetry of the solution, we obtain

$$\lim_{x\to 0^+}\frac{\partial}{\partial x}\begin{pmatrix}\phi\\ \psi_x\end{pmatrix}_P=\begin{pmatrix}-i\,\delta(z-z_0)\\ 0\end{pmatrix}. \tag{24}$$

From Eqs. (18)–(21), we obtain

$$(\lambda+2\mu)\left(u_x+\frac{\partial w}{\partial z}\right)=(\lambda+2\mu)\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial z^2}\right)=-\rho\omega^2\phi, \tag{25}$$

$$\mu\left(\frac{\partial u_x}{\partial z}-\frac{\partial^2 w}{\partial x^2}\right)=\mu\left(\frac{\partial^2\psi_x}{\partial x^2}+\frac{\partial^2\psi_x}{\partial z^2}\right)=-\rho\omega^2\psi_x. \tag{26}$$

From Eqs. (12) and (26), we obtain

$$(\lambda+2\mu)\frac{\partial^2 w}{\partial z^2}+\rho\omega^2 w+(\lambda+2\mu)\frac{\partial u_x}{\partial z}=-\rho\omega^2\psi_x. \tag{27}$$

Substituting Eqs. (25) and (27) into Eq. (24), we obtain

$$Q\frac{\partial}{\partial x}\begin{pmatrix}u_x\\ w\end{pmatrix}_P=\begin{pmatrix}i\rho\omega^2\delta(z-z_0)\\ 0\end{pmatrix}, \tag{28}$$

$$Q\equiv\begin{pmatrix}\lambda+2\mu & (\lambda+2\mu)\dfrac{\partial}{\partial z}\\ (\lambda+2\mu)\dfrac{\partial}{\partial z} & (\lambda+2\mu)\dfrac{\partial^2}{\partial z^2}+\rho\omega^2\end{pmatrix}. \tag{29}$$

Using Eq. (16) to replace the range derivative in Eq. (28) with a depth operator and performing row operations to eliminate Q, we obtain

$$(L^{-1}M)^{1/2}\begin{pmatrix}u_x\\ w\end{pmatrix}_P=\begin{pmatrix}k_p^2\delta(z-z_0)+\delta''(z-z_0)\\ -\delta'(z-z_0)\end{pmatrix}. \tag{30}$$

Applying Eq. (17) to march the field out a short distance in range from the source singularity, we obtain the self-starter,

$$\begin{pmatrix}u_x(x_0,z)\\ w(x_0,z)\end{pmatrix}_P=(L^{-1}M)^{-1/2}\exp(ix_0(L^{-1}M)^{1/2})\times\begin{pmatrix}k_p^2\delta(z-z_0)+\delta''(z-z_0)\\ -\delta'(z-z_0)\end{pmatrix}, \tag{31}$$

which can be implemented using a rational approximation and Galerkin's method. For the case of a shear source S at $z=z_0$, we similarly obtain

$$\begin{pmatrix}u_x(x_0,z)\\ w(x_0,z)\end{pmatrix}_S=\exp(ix_0(L^{-1}M)^{1/2})\begin{pmatrix}-\delta'(z-z_0)\\ \delta(z-z_0)\end{pmatrix}. \tag{32}$$

For the case of a point source in cylindrical geometry (r,z), we obtain

$$\begin{pmatrix}u_x(r_0,z)\\ w(r_0,z)\end{pmatrix}_P=r_0^{-1/2}(L^{-1}M)^{-1/4}\exp(ir_0(L^{-1}M)^{1/2})\times\begin{pmatrix}k_p^2\delta(z-z_0)+\delta''(z-z_0)\\ -\delta'(z-z_0)\end{pmatrix}, \tag{33}$$

$$\begin{pmatrix}u_x(r_0,z)\\ w(r_0,z)\end{pmatrix}_S=r_0^{-1/2}(L^{-1}M)^{1/4}\exp(ir_0(L^{-1}M)^{1/2})\times\begin{pmatrix}-\delta'(z-z_0)\\ \delta(z-z_0)\end{pmatrix}. \tag{34}$$

IV. GALERKIN'S METHOD AND INTERFACE CONDITIONS

In this section, we apply Galerkin's method to discretize L and M. Galerkin's method is based on the equations,

$$\int_0^H\left(\frac{\partial\sigma_{xx}}{\partial x}\eta-\sigma_{xz}\frac{d\eta}{dz}+\rho\omega^2 u\,\eta\right)dz=0, \tag{35}$$

$$\int_0^H\left(\frac{\partial\sigma_{xz}}{\partial x}\theta-\sigma_{zz}\frac{d\theta}{dz}+\rho\omega^2 w\,\theta\right)dz=0, \tag{36}$$

where $z=H$ is the bottom boundary of the computational grid and the arbitrary real test functions η and θ satisfy

$$\int_0^H\left(k_p^2\eta^2+\left(\frac{d\eta}{dz}\right)^2\right)dz<\infty, \tag{37}$$

$$\int_0^H\left(k_p^2\theta^2+\left(\frac{d\theta}{dz}\right)^2\right)dz<\infty. \tag{38}$$

Applying integration by parts to Eqs. (35) and (36), we obtain

$$\int_0^H\left(\frac{\partial\sigma_{xx}}{\partial x}+\frac{\partial\sigma_{xz}}{\partial z}+\rho\omega^2 u\right)\eta\,dz-\sigma_{xz}\eta|_0^H=0, \tag{39}$$

$$\int_0^H\left(\frac{\partial\sigma_{xz}}{\partial x}+\frac{\partial\sigma_{zz}}{\partial z}+\rho\omega^2 w\right)\theta\,dz-\sigma_{zz}\theta|_0^H=0. \tag{40}$$

We assume that the medium contains attenuation and that H is sufficiently large so that only weak signals reach the bottom boundary. We apply the boundary conditions $u_x=w=0$ at $z=H$ and also require that $\eta(H)=\theta(H)=0$. We do

not place constraints on η and θ at the surface $z=0$. Since η and θ are arbitrary, we conclude that the solution of Eqs. (39) and (40) satisfies Eqs. (1) and (2) for $0<z<H$. By considering test functions that vanish outside a small neighborhood of $z=0$, we conclude that the solution satisfies the boundary conditions $\sigma_{xz}=\sigma_{zz}=0$ at $z=0$. We determine the conditions at an interface by considering test functions that vanish outside a small neighborhood of the interface. If the displacements were not continuous, terms in Eqs. (39) and (40) would give rise to delta functions and their derivatives. We conclude that the displacements are continuous across the interface. By a similar argument, we conclude that σ_{xz} and σ_{zz} are continuous across the interface.

Difference formulas for the depth operators in L and M are obtained by choosing a set of simple test functions, approximating the solution in terms of these functions, substituting into Eqs. (35) and (36), and evaluating the integrals. We define the depth grid points $z_i=(i-1)\Delta z$, where Δz is the grid spacing, and work with piecewise linear test functions that vanish for $|z-z_i|>\Delta z$, increase linearly for $z_{i-1}<z<z_i$, and decrease linearly for $z_i<z<z_{i+1}$. We let α correspond to a coefficient and U correspond to a dependent variable. Substituting into Eqs. (35) and (36), we obtain the discretizations

$$\alpha U|_{z=z_i} \cong \frac{\alpha_{i-1}+\alpha_i}{12} U_{i-1} + \frac{\alpha_{i-1}+6\alpha_i+\alpha_{i+1}}{12} U_i + \frac{\alpha_i+\alpha_{i+1}}{12} U_{i+1}, \tag{41}$$

$$\alpha \frac{\partial U}{\partial z}\bigg|_{z=z_i} \cong -\frac{\alpha_{i-1}+2\alpha_i}{6\Delta z} U_{i-1} + \frac{\alpha_{i-1}-\alpha_{i+1}}{6\Delta z} U_i + \frac{2\alpha_i+\alpha_{i+1}}{6\Delta z} U_{i+1}, \tag{42}$$

$$\alpha \frac{\partial^2 U}{\partial z^2}\bigg|_{z=z_i} \cong \frac{\alpha_i}{(\Delta z)^2} U_{i-1} - \frac{2\alpha_i}{(\Delta z)^2} U_i + \frac{\alpha_i}{(\Delta z)^2} U_{i+1}, \tag{43}$$

$$\frac{\partial \alpha}{\partial z} U|_{z=z_i} \cong \frac{-\alpha_{i-1}+\alpha_i}{6\Delta z} U_{i-1} + \frac{-\alpha_{i-1}+\alpha_{i+1}}{6\Delta z} U_i + \frac{-\alpha_i+\alpha_{i+1}}{6\Delta z} U_{i+1}, \tag{44}$$

$$\frac{\partial \alpha}{\partial z}\frac{\partial U}{\partial z}\bigg|_{z=z_i} \cong \frac{\alpha_{i-1}-\alpha_i}{2(\Delta z)^2} U_{i-1} - \frac{\alpha_{i-1}-2\alpha_i+\alpha_{i+1}}{2(\Delta z)^2} U_i + \frac{-\alpha_i+\alpha_{i+1}}{2(\Delta z)^2} U_{i+1}. \tag{45}$$

Interface conditions must be explicitly enforced at a fluid-solid interface. We apply the (Δ,w) formulation in the fluid layer, where $p=-\lambda\Delta$ is the acoustic pressure. At an interface between fluid layer a and solid layer b, the conditions

$$\frac{\partial}{\partial z}(\lambda_a\Delta_a)+\rho_a\omega^2 w_b=0, \tag{46}$$

$$\lambda_a\Delta_a=\lambda_b u_{xb}+(\lambda_b+2\mu_b)\frac{\partial w_b}{\partial z}, \tag{47}$$

$$\frac{\partial}{\partial z}(\lambda_b u_{xb})+\frac{\partial}{\partial z}\left((\lambda_b+2\mu_b)\frac{\partial w_b}{\partial z}\right)+\rho_b\omega^2 w_b=0, \tag{48}$$

correspond to continuity of normal displacement, continuity of normal stress, and vanishing tangential stress. These conditions can be implemented using finite differences as described in Ref. 6.

One of the advantages of the combination of the (u_x,w) formulation and the Galerkin implementation is that the structure of the computational grid is convenient for problems involving sloping interfaces, which can be modeled with the stair-step approximation. Each rise corresponds to a vertical interface, which is placed midway between two points in the range grid. The Galerkin solution places each run midway between two points in the depth grid, and there is a one-to-one correspondence between grid points across a vertical interface. With this grid structure, it is straightforward to conserve the dependent variables across a vertical interface (improved accuracy may be achieved by conserving other quantities, but we do not address this issue here). The (Δ,w) formulation requires the explicit implementation of second-order interface conditions on the runs. Various numerical solutions are possible, but a standard centered-difference scheme involves a double grid point placed on a run. Two values of the solution (which is not continuous across the interface) are assigned to this grid point, one value for each side of the horizontal interface. It is not clear how to handle a vertical interface, where a double grid point on one side of the interface must somehow be converted into a single grid point on the other side of the interface and vice versa.

V. EXAMPLES

In this section, we compare parabolic equation solutions with a reference solution that is based on separation of variables.[23] We show that the (u_x,w) solution is accurate for problems involving solid-solid interfaces, Rayleigh waves, and Stoneley waves. We consider problems in cylindrical geometry and use Eq. (33).

Example A illustrates the ability of the (u_x,w) solution to handle solid-solid interfaces. A 50-Hz source is placed at $z=98$ m in a 100-m water column in which the sound speed is 1500 m/s. A 10-m sediment layer overlies a half-space basement. The sediment parameters are $c_p=1700$ m/s, $c_s=800$ m/s, $\rho=1.3$ g/cm^3, $\beta_p=0.1$ dB/λ, and $\beta_s=0.2$ dB/λ. The basement parameters are $c_p=2400$ m/s, $c_s=1200$ m/s, $\rho=1.7$ g/cm^3, $\beta_p=0.1$ dB/λ, and $\beta_s=0.2$ dB/λ. In Fig. 1, the (Δ,w) and (u_x,w) solutions (both implemented with Galerkin's method) are compared with the reference solution. Since Δ is not continuous across interfaces, delta functions may arise when the (Δ,w) solution is implemented with Galerkin's method. The (Δ,w) solution is accurate for some problems,[14] but it breaks down for example A. The

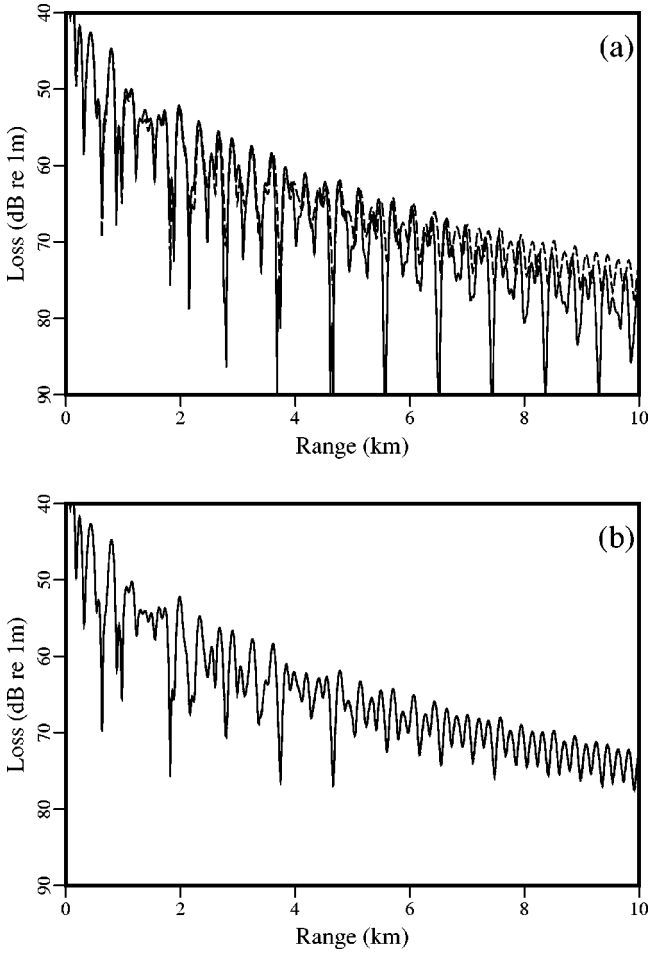


FIG. 1. Transmission loss at $z=60$ m for example A, which involves an elastic sediment layer overlying an elastic half-space. The dashed curves are the reference solution. The solid curves are (a) the (Δ,w) solution, which breaks down, and (b) the (u_x,w) solution, which is accurate.

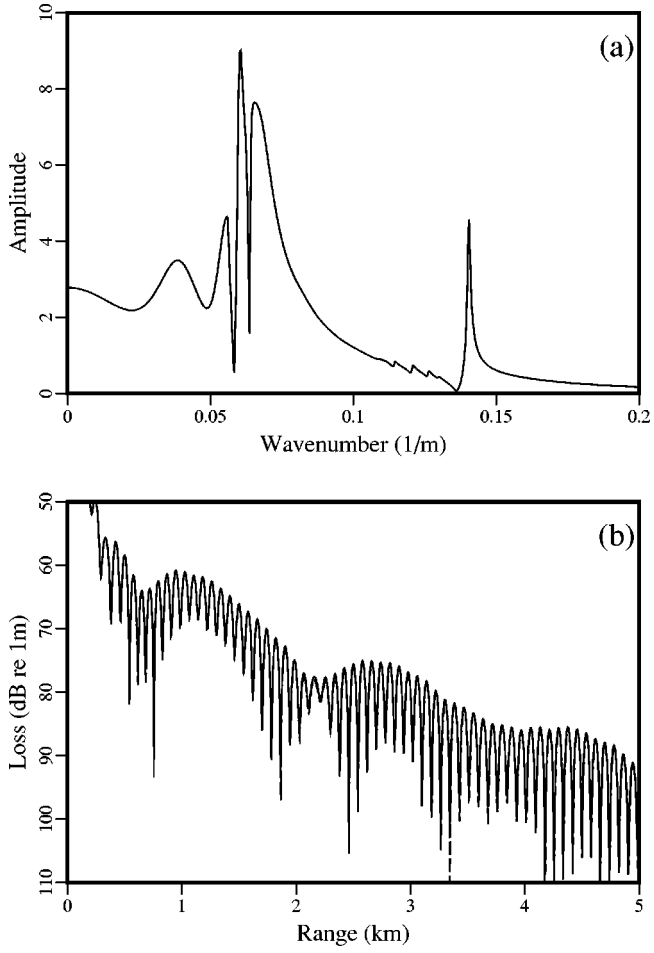


FIG. 3. Compressional wave energy at $z=20$ m for example C, which involves an elastic layer overlying an elastic half-space. (a) The peak to the right in the wave-number spectrum corresponds to a Rayleigh wave at the free surface. (b) The (u_x,w) solution (solid curve) is in agreement with the reference solution (dashed curve).

(u_x,w) solution is accurate for example A. The (u_x,w) solution was obtained using $\Delta z=0.0625$ m, which corresponds to 256 grid points per wavelength at 800 m/s. To achieve similar accuracy for an acoustics problem,[24] it was necessary to use 120 grid points per wavelength. For example A, it is necessary to use finer sampling in order to handle the nearly vertical propagation of shear waves in the sediment.

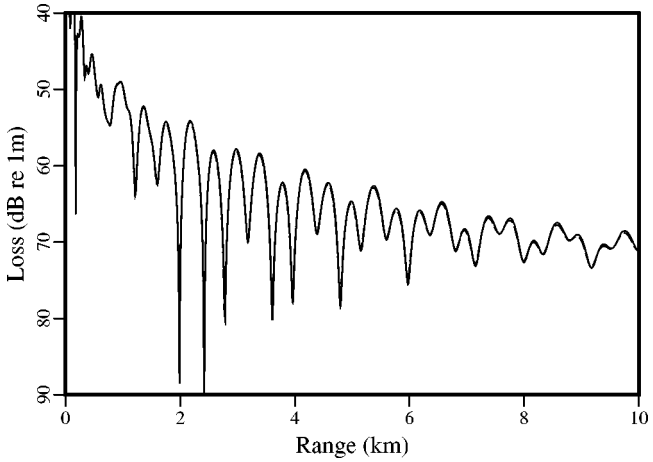


FIG. 2. Transmission loss at $z=60$ m for example B, which involves a low-speed elastic sediment layer overlying an elastic half-space. The (u_x,w) solution (solid curve) is in agreement with the reference solution (dashed curve).

Example B illustrates the ability of the (u_x,w) solution to handle a layer with low shear speed. A 50-Hz source is placed at $z=98$ m in a 100-m water column in which the sound speed is 1500 m/s. A 20-m sediment layer overlies a half-space basement. The sediment parameters are $c_p=1700$ m/s, $c_s=100$ m/s, $\rho=1.2$ g/cm^3, $\beta_p=0.5$ dB/λ, and $\beta_s=0.5$ dB/λ. The basement parameters are $c_p=2400$ m/s, $c_s=1200$ m/s, $\rho=1.5$ g/cm^3, $\beta_p=0.5$ dB/λ, and $\beta_s=0.5$ dB/λ. In Fig. 2, the (u_x,w) solution is in agreement with the reference solution. Despite the low shear speed in the sediment layer, an accurate solution was obtained using $\Delta z=0.0625$ m.

Example C illustrates the ability of the (u_x,w) solution to handle Rayleigh waves. A 25-Hz source is placed at $z=10$ m in a 200-m solid layer that overlies a solid half-space. The parameters in the layer are $c_p=2400$ m/s, $c_s=1200$ m/s, $\rho=1.5$ g/cm^3, $\beta_p=0.1$ dB/λ, and $\beta_s=0.2$ dB/λ. The parameters in the half-space are $c_p=2800$ m/s, $c_s=1400$ m/s, $\rho=1.7$ g/cm^3, $\beta_p=0.1$ dB/λ, and $\beta_s=0.2$ dB/λ. The Rayleigh wave corresponds to the sharp

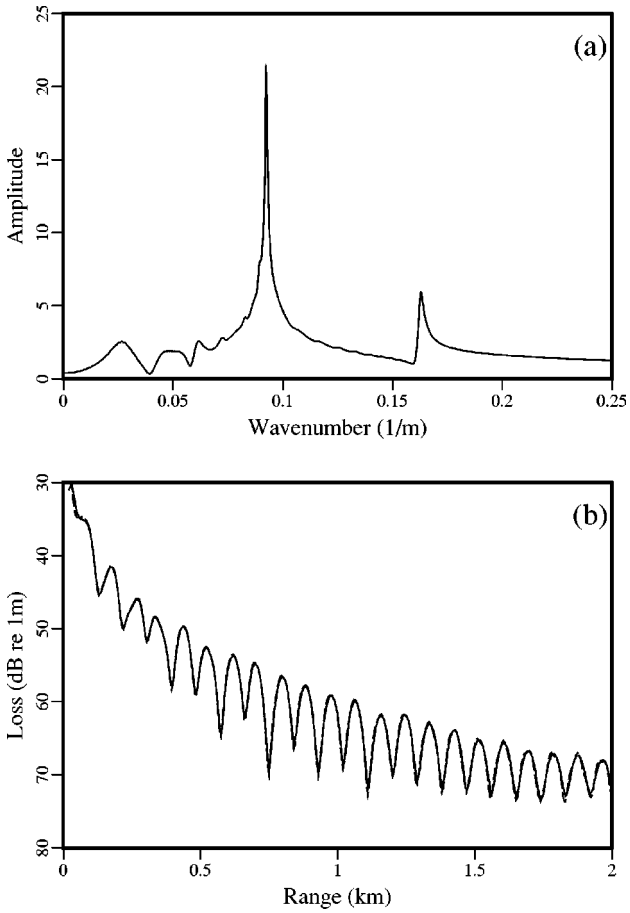


FIG. 4. Compressional wave energy at $z=299$ m for example D, which involves an elastic layer overlying an elastic half-space. (a) The peak to the right in the wave-number spectrum corresponds to a Stoneley wave at the solid-solid interface. (b) The (u_x,w) solution (solid curve) is in agreement with the reference solution (dashed curve).

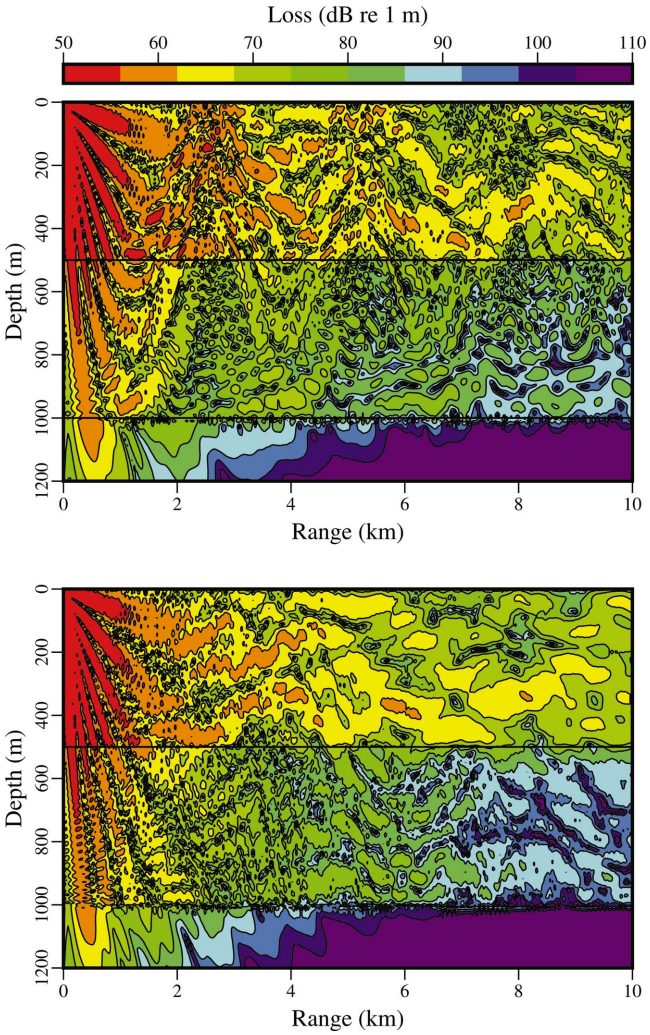


FIG. 5. Compressional wave energy for example E, which involves a refracting sediment layer overlying a homogeneous basement (top frame). Evidence of refraction can be seen in the sediment layer. There is no evidence of refraction for the case of a homogeneous sediment (bottom frame).

peak to the right in the wave-number spectrum appearing in Fig. 3. The (u_x,w) solution appearing in Fig. 3 is in agreement with the reference solution.

Example D illustrates the ability of the (u_x,w) solution to handle Stoneley waves. A 25-Hz source is placed at $z=295$ m in a 300-m solid layer that overlies a solid half-space. We select the parameters in order to magnify the effect of the Stoneley wave, which exists when $\lambda=\mu$ and there is a large density contrast but small shear speed contrast across the interface.[25] In the layer and half-space, we take $c_p=1700$ m/s, $c_s=981.5$ m/s, $\beta_p=0.2$ dB/λ, and $\beta_s=0.4$ dB/λ. We take $\rho=1.2$ g/cm^3 in the layer and $\rho=3.6$ g/cm^3 in the half-space. The Stoneley wave corresponds to the peak to the right in the wave-number spectrum appearing in Fig. 4. The (u_x,w) solution appearing in Fig. 4 is in agreement with the reference solution.

Example E illustrates the effects of gradients in the wave speeds. A 35-Hz source is placed at $z=150$ m in a 500-m water column in which the sound speed is 1500 m/s. A refracting sediment layer that is 500 m thick overlies a homogeneous basement. The wave speeds depend linearly on z in the sediment layer, with $c_p=1600$ m/s and $c_s=700$ m/s at the top of the layer and $c_p=2400$ m/s and $c_s=1200$ m/s at the bottom of the layer. The other sediment parameters are $\rho=1.2$ g/cm^3, $\beta_p=0.1$ dB/λ, and $\beta_s=0.2$ dB/λ. The basement parameters are $c_p=3400$ m/s, $c_s=1700$ m/s, $\rho=1.5$ g/cm^3, $\beta_p=0.25$ dB/λ, and $\beta_s=0.5$ dB/λ. The (u_x,w) solution is illustrated in Fig. 5. Some of the beams of the Lloyd's mirror pattern near the source penetrate into the sediment and get refracted back toward the water column. Also appearing in Fig. 5 is the solution for the case in which the sediment layer is homogeneous, with $c_p=1600$ m/s, $c_s=700$ m/s, $\rho=1.2$ g/cm^3, $\beta_p=0.1$ dB/λ, and $\beta_s=0.2$ dB/λ. The beams are not refracted for this case, and there are significant differences in the water column.

VI. CONCLUSION

The elastic parabolic equation has been implemented in terms of the dependent variables (u_x,w). This formulation is better suited to handling solid-solid interfaces. The (u_x,w) solution can be implemented using Galerkin's method and accurately handles all types of interface waves. The Galerkin implementation of the (Δ,w) solution breaks down for some problems involving solid-solid interfaces. The Galerkin implementation of the (u_x,w) solution is expected to lead to improvements in the treatment of problems that involve slop-

ing interfaces, which are difficult to handle in the (Δ, w) formulation. The self-starter was generalized to the (u_x, w) case for both compressional and shear sources.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research.

[1] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994), pp. 343–412.
[2] R. H. Hardin and F. D. Tappert, "Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev. **15**, 423 (1973).
[3] S. M. Flatté and F. D. Tappert, "Calculation of the effect of internal waves on oceanic sound transmission," J. Acoust. Soc. Am. **58**, 1151–1159 (1975).
[4] F. D. Tappert, "The parabolic approximation method," in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis, Lecture Notes in Physics, Vol. 70 (Springer, New York, 1977), pp. 224–287.
[5] R. R. Greene, "A high-angle one-way wave equation for seismic wave propagation along rough and sloping interfaces," J. Acoust. Soc. Am. **77**, 1991–1998 (1985).
[6] M. D. Collins, "A higher-order parabolic equation for wave propagation in an ocean overlying an elastic bottom," J. Acoust. Soc. Am. **86**, 1459–1464 (1989).
[7] B. T. R. Wetton and G. H. Brooke, "One-way wave equations for seismoacoustic propagation in elastic waveguides," J. Acoust. Soc. Am. **87**, 624–632 (1990).
[8] M. D. Collins, "Higher-order parabolic approximations for accurate and stable elastic parabolic equations with application to interface wave propagation," J. Acoust. Soc. Am. **89**, 1050–1057 (1991).
[9] M. B. Porter, F. B. Jensen, and C. M. Ferla, "The problem of energy conservation in one-way models," J. Acoust. Soc. Am. **89**, 1058–1067 (1991).
[10] M. D. Collins and E. K. Westwood, "A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed, and density," J. Acoust. Soc. Am. **89**, 1068–1075 (1991).
[11] M. D. Collins, "An energy-conserving parabolic equation for elastic media," J. Acoust. Soc. Am. **94**, 975–982 (1993).
[12] M. D. Collins and W. L. Siegmann, "A complete energy conservation correction for the elastic parabolic equation," J. Acoust. Soc. Am. **105**, 687–692 (1999).
[13] M. D. Collins and R. B. Evans, "A two-way parabolic equation for acoustic back scattering in the ocean," J. Acoust. Soc. Am. **91**, 1357–1368 (1992).
[14] M. D. Collins, "A two-way parabolic equation for elastic media," J. Acoust. Soc. Am. **93**, 1815–1825 (1993).
[15] J. F. Lingevitch and M. D. Collins, "Wave propagation in range-dependent poro-acoustic waveguides," J. Acoust. Soc. Am. **104**, 783–790 (1998).
[16] M. D. Collins and D. K. Dacol, "A mapping approach for handling sloping interfaces," J. Acoust. Soc. Am. **107**, 1937–1942 (2000).
[17] A. J. Fredricks, W. L. Siegmann, and M. D. Collins, "A parabolic equation for anisotropic elastic media," Wave Motion **31**, 139–146 (2000).
[18] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method* (Prentice-Hall, Englewood Cliffs, NJ, 1973).
[19] H. Kolsky, *Stress Waves in Solids* (Dover, New York, 1963).
[20] M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[21] M. D. Collins, "A self-starter for the parabolic equation method," J. Acoust. Soc. Am. **92**, 2069–2074 (1992).
[22] M. D. Collins, "The stabilized self starter," J. Acoust. Soc. Am. **106**, 1724–1726 (1999).
[23] Reference 1, pp. 203–270.
[24] F. B. Jensen and C. M. Ferla, "Numerical solutions of range-dependent benchmark problems," J. Acoust. Soc. Am. **87**, 1499–1510 (1990).
[25] J. Miklowitz, *The Theory of Elastic Waves and Waveguides* (North-Holland, New York, 1978), pp. 165–168.

Acoustic backscatter measurements from littoral seabeds at shallow grazing angles at 4 and 8 kHz

Paul C. Hines, John C. Osler, and Darcy J. MacDougald
Defence R&D Canada—Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 28 April 2004; revised 4 January 2005; accepted 4 March 2005)

Direct measurement of acoustic scattering from the seabed at shallow grazing angles and low kilohertz frequencies presents a considerable challenge in littoral waters. Specifically, returns from the air–water interface typically contaminate the signals of interest. To address this issue, DRDC Atlantic has developed a sea-going research system for measuring acoustic scatter from the seabed in shallow-water environs. The system, known as the wideband sonar (WBS), consists of a parametric array transmitter and a superdirective receiver. In this paper, backscatter measurements obtained with the WBS at two sandy, shallow-water sites off North America's Atlantic coast are presented. Data were collected at 4 and 8 kHz at grazing angles from 3°–15°. The backscattering strength is similar at both sites and, below about 8°, it appears to be independent of frequency within the statistical accuracy of the data. The measurements show reasonable agreement with model estimates of backscatter from sandy sediments. A small data set was collected at one of the sites to examine the feasibility of using the WBS to measure the azimuthal variability of acoustic scatter. The data set—although limited—indicates that the parametric array's narrow beamwidth makes the system well-suited to this task. [DOI: 10.1121/1.1898064]

PACS numbers: 43.30.Gv, 43.30.Ft, 43.30.Hw [RAS] Pages: 3504–3516

I. INTRODUCTION

Quantifying the effect that the seabed has on an acoustic wave is crucial if one is to assess the performance of current sonar systems in shallow waters. Furthermore, advances in sonar technology require a clear understanding of how sound scatters from the seabed, which in turn requires experiments to complement theoretical and numerical models. The measurements can be classified as direct or reverberation. The former refers to measurements that have a single interaction with a boundary, whereas the latter measurements encompass returns from many boundary interactions that are time coincident at the receiver. Although reverberation experiments provide a reasonable approximation to naval systems,[1] the measurements are averaged over many grazing angles and boundaries, and can obscure the physical mechanisms that provide the key to understanding the physics of scattering. In recent years a number of direct measurements of backscattering have been made in shallow water at frequencies above 10 kHz (see Williams *et al.*[2] and the references therein). However, in shallow coastal water at low kilohertz frequencies (nominally below 10 kHz), direct measurements are particularly challenging. This is because the size of a conventional sonar is roughly proportional to its acoustic wavelength and, in order to project a narrow-enough acoustic beam, the sonar must be many wavelengths long. In this circumstance, one usually cannot place the sonar near enough to the seabed to avoid vertically directed sidelobes which generate unwelcome interference from sea-surface scatter. Because of these challenges, and the fairly recent focus on shallow-water acoustics in this frequency range, the literature containing direct measurements of scattering below 10 kHz is rather limited. Bunchuk and Zhitkovskii[3] present backscattering data from 2–16 kHz collected in 250 m of water, but it is unclear from the text whether the data should be classified as reverberation or direct measurements. Holland[4] collected direct measurements of backscattering data from 1–4 kHz by deploying a vertical line array from a drifting ship. Placing the source at the base of the array allowed him to separate scatter from the seabed and the sea surface. The approach is describe in detail in Ref. 5. Soukup and Gragg[6] present backscatter measurements taken over a limestone seabed in the 2–3.5-kHz region. Several papers contained in Ref. 7 also present data of direct measurements of scattering. Day and Yamamoto[8] present backscatter data collected at 3.75, 7.5, and 15 kHz in 17-m waters near Ft. Pierce, FL. Given the variability in composition of coastal seabeds, existing data sets are still rather limited.

To address this deficiency, Defense R&D Canada—Atlantic (formerly DREA) has developed a wideband active sonar to interrogate the seabed in shallow water and quantify its geoacoustic properties. During operation, the system (referred to as WBS) is fully decoupled from the tending research vessel and is instrumented with both acoustic and nonacoustic sensors to assist in the evaluation of the data. In May 2001 the system was deployed as part of a joint US-SACLANTCEN-CA experiment referred to as Boundary Characterization 2001. The WBS system collected acoustic backscatter data as a function of azimuth and grazing angle at two shallow-water sites off the east coast of North America—one near New Jersey and one near Nova Scotia. Measurements were made at grazing angles ranging from about 3° up to 15°, at frequencies of 4 and 8 kHz.

Following the introduction, an overview of the WBS system and its technical functionality are given. Then, the experimental geometry and methodology is described. Finally, results from the experiments are presented with a primary focus on the estimates of backscattering strength obtained.

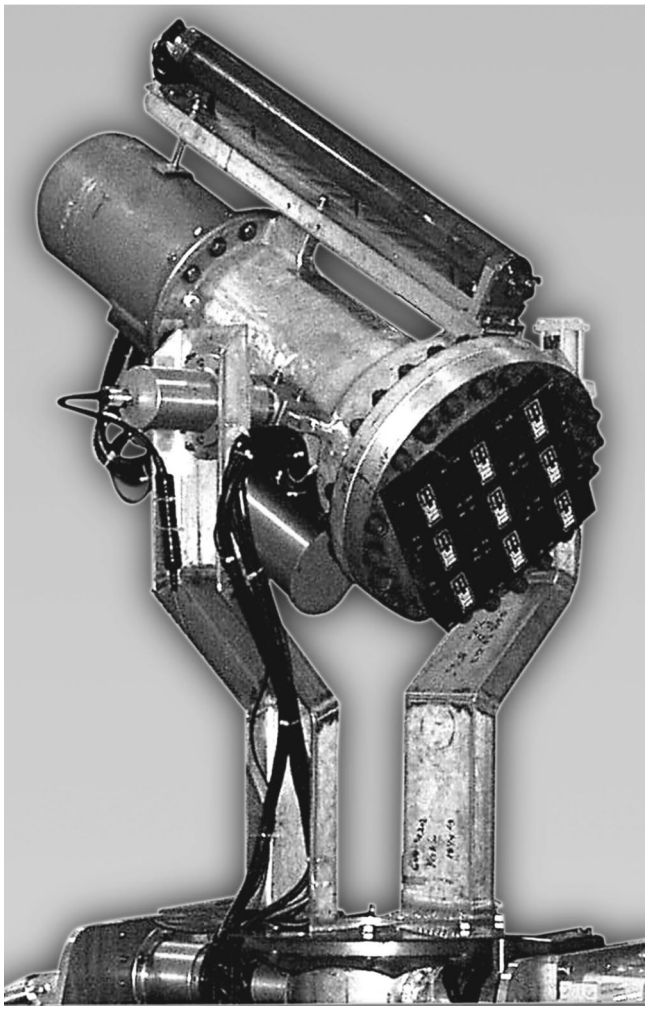


FIG. 1. Photograph of the wideband sonar (WBS) armature yoke holding the parametric array transmitter system (PATS). The nine elements that comprise the parametric source are visible on the right-hand side of the photo. The SIREM line array is shown mounted along the length of the PATS housing. The array hydrophones are hidden from view by the polyurethane housing that surrounds them.

II. THE WIDEBAND SONAR (WBS)

Defense R&D Canada-Atlantic (DRDC Atlantic) has developed[9,10] a wideband sonar system for collecting environmental acoustic data in the open ocean. The WBS is remotely controlled from a research ship via an rf radio link fixed to the system's surface float. This minimizes the risk of acoustic interference from the ship and prevents ship motion from compromising its stability. The acoustic sensor suite consists of a parametric array[11,12] transmitter (PATS), a superdirective endfire line array receiver (SIREM), and a triaxial intensity array receiver (SIRA). In addition to acoustic sensors, the array is instrumented with an assortment of nonacoustic sensors to assist in evaluating the data. The nonacoustic sensors include depth, tilt, roll, and heading sensors to monitor array position and direction, as well as accelerometers to monitor platform vibration. Whereas SIRA is fixed to the space frame, the sonar armature which holds PATS and SIREM can be panned through 360° in azimuth (see Fig. 1). Pointing direction is monitored using the nonacoustic sensors mounted on the space frame combined with optical shaft encoders to measure the pan and tilt angle of the sonar head. This obtains a resolution of approximately $\pm 1°$ in each of the vertical and horizontal planes. The system can be configured to be either bottom tethered (Fig. 2 left) or bottom mounted (Fig. 2 right). In bottom-tethered mode the armature rotates 180° vertically so that the sonar can be positioned above or below the space frame. This enables measurements through 4π steradians. Platform stability is achieved by decoupling the space frame from the surface float through a weighted cable and streaming the space frame into the prevailing shear current. In bottom-mounted mode a remote command is sent from the ship to the surface float to flood the subsurface floats on the space frame to set the system on the seabed. This offers an extremely stable platform that permits coherent averaging of multiple pings. This is used in low-SNR conditions such as measurements of backscattering strength at very shallow grazing angles; however, physical constraints limit the vertical range of angles to $-30°$ from the horizontal up to $+90°$. In bottom-mounted mode the sonar head sits 1.5 m above the space frame, which corresponds to 2.7 m above the seabed. Prior to recovery compressed air is forced into the subsurface floats to evacuate the water, thereby leaving the space frame positively

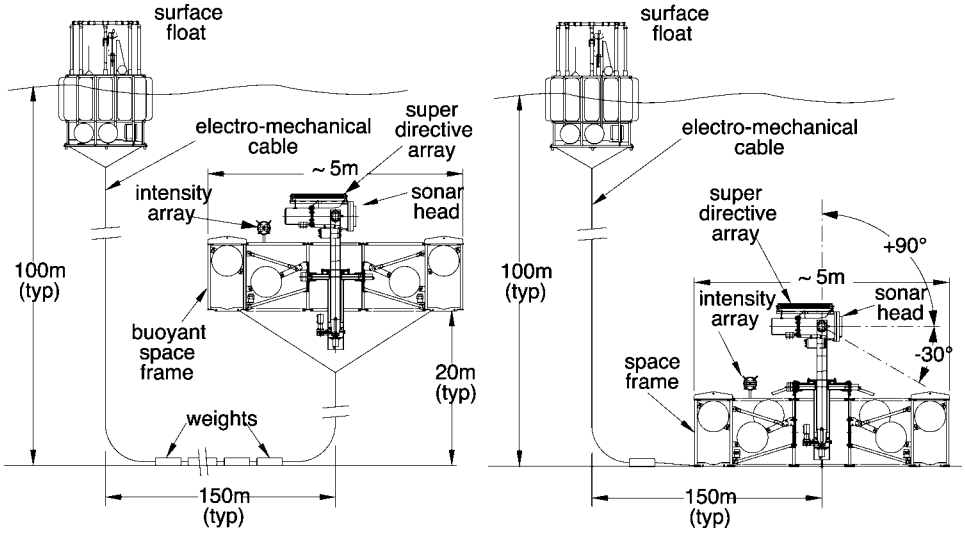


FIG. 2. Schematic of WBS bottom tethered (left) and bottom mounted (right). Note that the pivot point for the sonar support arm is different for the two configurations.

buoyant so that it rises to the sea surface. The experiments discussed in this document were all performed with the system in bottom-mounted mode.

Employing a parametric source offers three advantages: First and foremost, due to the nature of signal generation in the parametric array, no sidelobes are formed. This feature avoids the added complexities of unwanted boundary interactions when making measurements in shallow water. This enables one to measure, for example, the reflection and scattering characteristics of the seabed in the absence of interference from sonar sidelobes. Second, one can obtain a wider bandwidth than that obtained using a conventional source. In the present case a bandwidth of 2 to 10 kHz is realized from a single transducer with a primary frequency of 100 kHz. The wide bandwidth means that a single transducer can be used in place of a suite of transducers, each of which may require separate power and tuning circuitry. The wide bandwidth also allows a great deal of flexibility in the types of pulses available. Third, the beamwidth of the parametric array is extremely narrow relative to the transmitter aperture. In the present case a square transducer measuring 41 cm on a side yielded horizontal and vertical beamwidths of approximately 4°–7°. This facilitates measurements of backscatter as a function of azimuth as well as out-of-plane bistatic scatter. The price paid for these advantages is poor acoustic efficiency, $<0.1\%$ across the difference frequency band.

The principle receiver for measuring acoustic scatter is SIREM—a six-channel superdirective hydrophone line array.[13] Superdirective arrays compute pressure gradients and as such require interelement spacings that are a small fraction of an acoustic wavelength. Thus, by its very nature, the superdirective array is much more compact than a conventional array and complements the compactness of the parametric transmitter. The array yields gains of up to 15 dB across the sonar's frequency band from an array aperture which is only 0.8 m long. The SIRA intensity array[14] is a secondary receiver and is used to localize the platform and to measure ambient noise directionality. All backscattering strength estimates results reported in this paper were obtained using a single hydrophone from the SIREM line array as the receiver.

The WBS system was used to collect acoustic backscatter data at two shallow-water locations off the eastern coast of North America. Prior to reporting on these results, the source characteristics of the parametric array are presented in the following section.

III. PARAMETRIC ARRAY CALIBRATION

The calibration experiments were performed at DRDC Atlantic's calibration barge, located in Bedford Basin, in Halifax Harbor. The barge is an enclosed, 300-metric ton steel vessel, moored in a 42-m-deep, salt-water basin. The barge measures 36×17 m. Inside the barge superstructure is a rectangular well (18×9 m) open to the sea (sound speed $c\approx1460$ m/s). Two mobile trolleys, which are equipped with mounting stations capable of handling up to 140 kg, traverse the well. The trolley/mounting-station configuration allows the sensors to be located anywhere within the well. A larger variable-depth station capable of handling up to 7000 kg is located at the northern end of the well equidistant from the sides. There is a second variable-depth station capable of handling up to 7000 kg located 30 m away at the southern end of the barge outside the superstructure. This latter station allows the maximum separation between source and receiver. Figure 3 shows a schematic of the facility and Fig. 4 shows a photograph of its interior.

The source level and beamwidth of a parametric array is range dependent in the interaction zone.[12] Therefore, the calibrations were performed at several ranges to obtain the source level and beamwidth at ranges corresponding to the near field and the far field. First, the PATS (recall Fig. 1) was mounted to the large outside station and set to a depth of approximately 6 m. A 2.54-cm spherical calibration hydrophone (sensitivity -195 re: 1 V/μPa) was mounted to one of the small trolley stations and lowered to the same depth. The geometry is shown in Fig. 3. The hydrophone was frequency independent and omnidirectional to within ±1 dB across the difference frequency band of 2–10 kHz. The hydrophone had a first resonance at 54 kHz, above which its sensitivity decreased monotonically up to 150 kHz. The receive sensitivity at 100 kHz was -217 dB re: 1 V/μPa; that is, -22 dB relative to its response across the difference frequency band. This lower sensitivity at 100 kHz was chosen deliberately to ensure that the excessive source level at the primary frequency (nominally 240 dB re: 1 μPa@1 m) did not generate a nonlinear (difference frequency) response in the hydrophone ceramic and contaminate the true difference frequency signal generated within the water column.[15] The on-axis sound-pressure level (SPL) and the beam pattern were measured at difference frequencies of $f_d=2$, 3, 4, 5, 8, and 10 kHz at separations ranging from 15–30 m. This corresponds to the minimum and maximum separations for this geometry. Next, the PATS was mounted to the interior station and the measurements were repeated at ranges of 3–15 m. Time constraints required that these close-range measurements were limited to 2, 4, and 8 kHz. For each calibration, a series of pulses of 2-ms duration was collected, converted to power spectra, and averaged. Each data point plotted in the calibration figures shown hereafter represents the power average of no fewer than 16 pings.

To examine the impact of nonlinearity in the calibration hydrophone, the parametric array was truncated using an acoustic filter and a set of on-axis source-level measurements was made. The acoustic filter absorbs the primary frequency (nominally 100 kHz) while allowing the difference frequency to pass virtually unattenuated. Since the effect of hydrophone nonlinearity is most pronounced at close range and low difference frequencies,[15] this experiment was performed at $f_d=2$ kHz at ranges of 3–15 m.

Figure 5 shows the SPL data converted to on-axis difference-frequency source level (SL_d) referenced to 1 m, as a function of range for frequencies of 2, 4, and 8 kHz. The SL_d at 2 kHz for the truncated array is also shown. The increase in the SL_d estimate for ranges less than 10 m is clearly visible in the 2-kHz data and results from the nonlinear response of the calibration hydrophone. This contamination can be seen to a lesser degree in the 4-kHz data as well. From 10 to 15 m there is remarkably good agreement between the truncated and untruncated 2-kHz data, suggesting

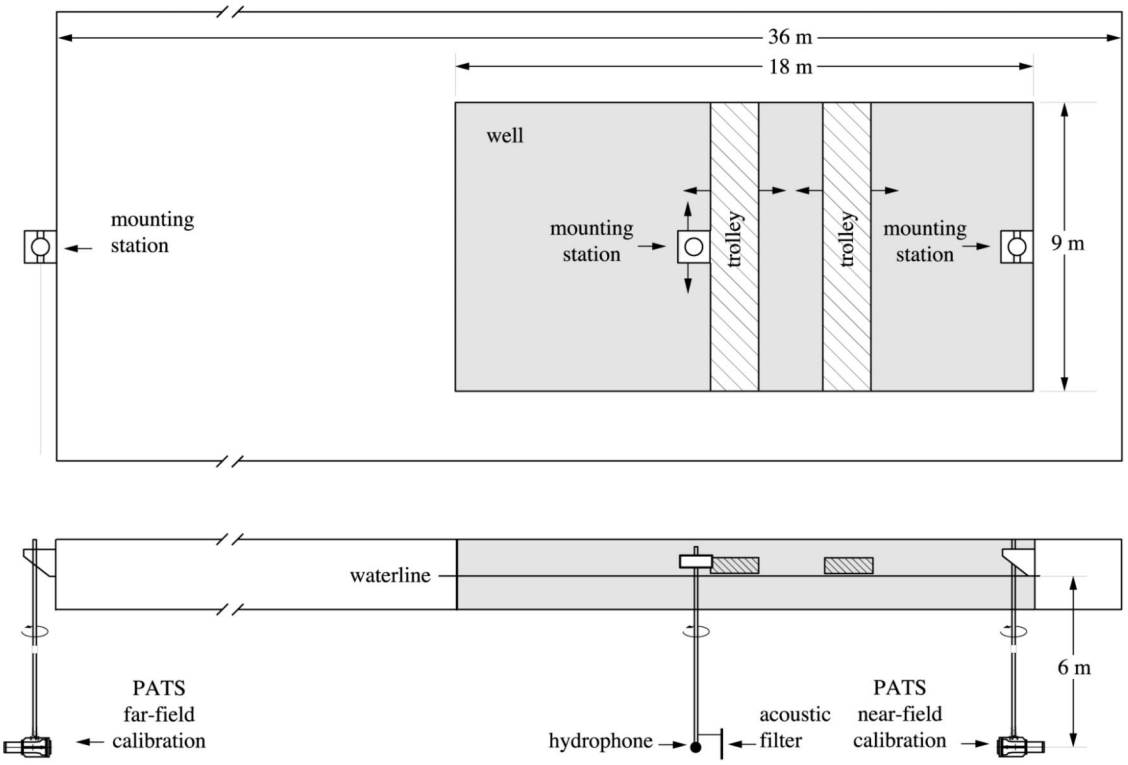


FIG. 3. Schematic of parametric array calibration experiment showing a plan view (top) and cross-section view (bottom). The superstructure surrounding the well is not shown.

that the SL_d estimate obtained using the truncated array is valid at closer ranges. To support this view, as well as to estimate the SL_d at close range at higher frequencies, a numerical model[16] was evaluated to estimate the source level in the near field of the parametric array. Figure 6 compares the model results to the data from Fig. 5. The agreement between the model and the data at 2 kHz is excellent. For ranges greater than about 10 m the model shows very good agreement with the data at 4 and 8 kHz. At ranges less than 10 m disagreement between the model and data can be attributed to hydrophone nonlinearity. Figure 7 summarizes the range calibration results at 2, 3, 4, 5, and 8 kHz and the corresponding near-field model estimates.

The beam pattern measured at a difference frequency of 8 kHz is plotted in Fig. 8 for ranges of 10–30 m. Clearly the beam pattern is less sensitive to range than is the source level. The data exhibit a monotonic decrease in amplitude off-axis typical of a parametric array, with no sidelobes evident out to at least 50°. Figure 9 contains the beam pattern measurements at frequencies of 2, 4, and 8 kHz taken at a

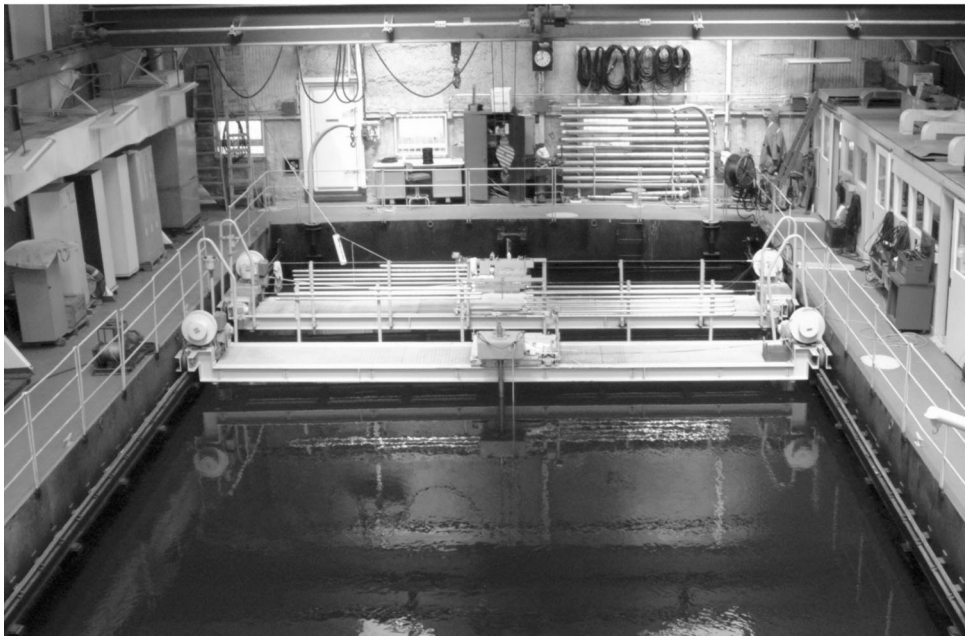


FIG. 4. Photograph of the interior of DRDC Atlantic's calibration barge. The mounting trolleys can be seen in the far end of the well.

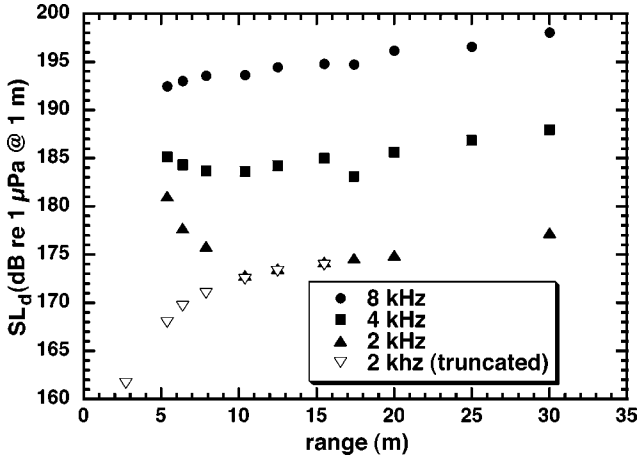


FIG. 5. Calibration results showing difference-frequency source level (SL_d) as a function of range for the truncated and untruncated parametric array.

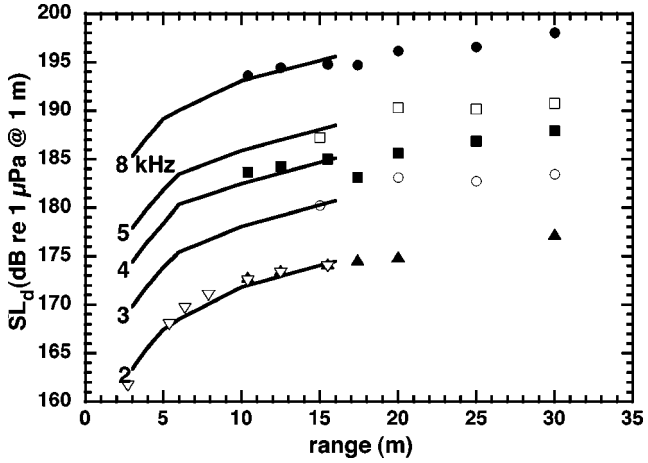


FIG. 7. Summary of the SL_d data and model results as a function of range for 2, 3, 4, 5, and 8 kHz. The units are only displayed at 8 kHz to prevent overcrowding the figure.

range of 30 m. The results obtained at 3 and 5 kHz are consistent with Fig. 9 but have been omitted from the figure to avoid overcrowding.

IV. THE EXPERIMENTAL SITES

The experiments were performed at two shallow-water sites on North America's eastern seaboard—one site known as ONR Strataform East is located off the New Jersey coast and the second site is about 200 km southeast of Halifax, Nova Scotia on Canada's Scotian Shelf. The sites were chosen primarily because of the availability of supporting geotechnical measurements to assist in data interpretation. References 17 and 18 describe the locations in detail; however, in an effort to keep this paper self-contained, a brief description of each site is given below.

A. Strataform East site

Site NJ1 is located on the continental shelf off the coast of New Jersey, in a ''natural laboratory'' established by the Office of Naval Research (ONR) (See Fig. 10.) Numerous complementary acoustic and geophysical experiments and surveys have been conducted in the ONR Strataform East study area. Site NJ1 on the New Jersey Shelf is located on a midshelf sand ridge, oriented approximately 060 °T, with an rms height of approximately 1 to 1.5 m, and width of approximately 1.5 to 2 km.[17] The slope of the seabed within 100 m of the WBS (that is, grazing angles down to 1.5°) in the directions of the scattering experiments is less than 0.04°. The sand ridges in this area appear to be relic features on which erosion and sediment transport processes are not active; the erosional processes are restricted to the adjacent swale areas characterized by ribbons and scour pits. Although the swath bathymetry data in Ref. 17 are not capable of resolving higher frequency ripples, there are no known observations of ripples being found on the midshelf sand ridges. Extensive seismic reflection profiling has revealed that this area is underlain by a NNW–SSE-oriented dendritic drainage system. In the vicinity of NJ1, there is a series of ancient river channels that are approximately 10 m deep and have been filled with unconsolidated sediments.[19] These river channels and a prominent regional near-surface seismic reflector, the *R* reflector, have been identified as potential sources of coherent energy scattered from the seabed that may manifest itself as false targets on active sonar systems.

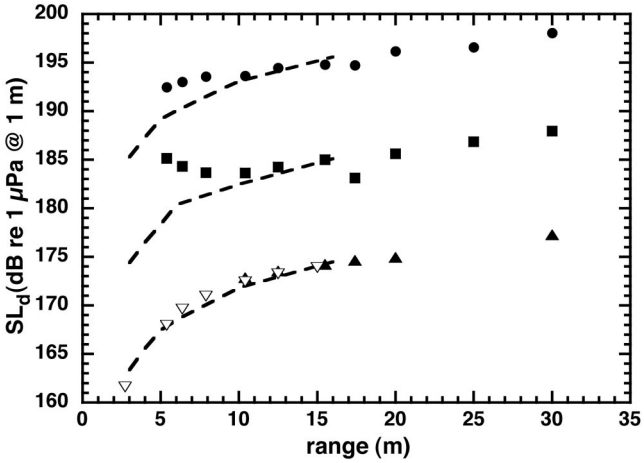


FIG. 6. Model-data comparison for SL_d calibration of parametric array at 2 kHz (triangles), 4 kHz (squares), and 8 kHz (circles). The corresponding model results are given by the dashed lines.

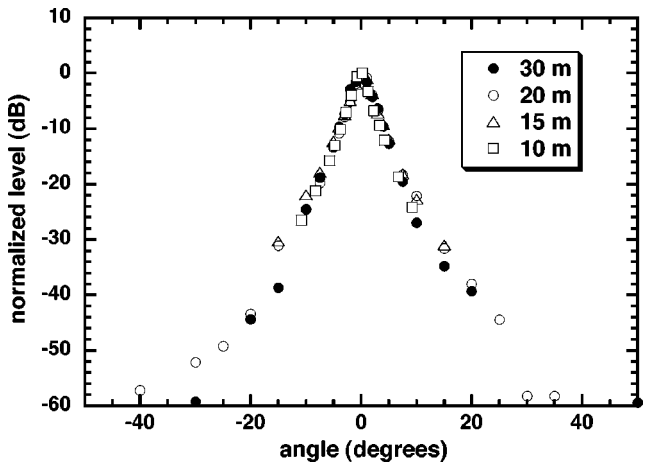


FIG. 8. Parametric array beam pattern calibration results at 8 kHz for ranges from 10–30 m. The beam pattern is insensitive to range within the region measured.

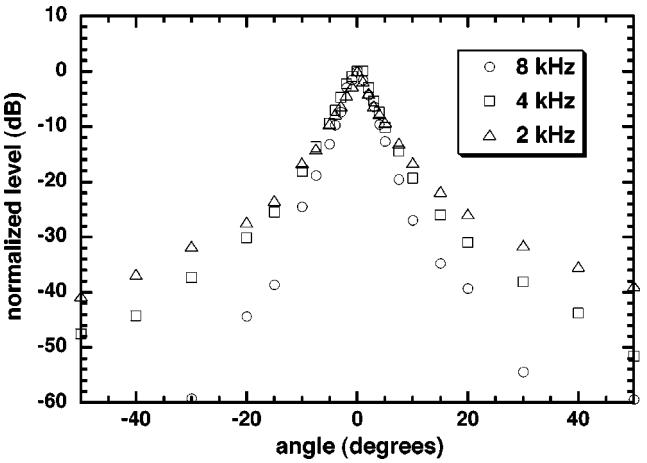


FIG. 9. Parametric array beam pattern measured at 30-m range at frequencies of 2, 4, and 8 kHz. The results obtained at 3 and 5 kHz are consistent with those shown but have been omitted from the figure to avoid overcrowding.

The surficial sediment at NJ1 is a well-sorted medium sand,[20] grain size of 0.25 to 0.35 mm (ϕ=1.5 to 2.0). The average compressional wave speed measured at 65 kHz at two nearby locations using the University of New Hampshire ISAAP sediment probe[21] is 1727 and 1756 ms^{-1}.

B. Scotian Shelf site

The surficial sediment distribution of the Scotian Shelf is closely related to the processes that transpired during and since the last glaciation. In general, the shallow-water bank areas tend to have sand or gravel overlying glacial till and then a consolidated rock layer (sedimentary offshore and metamorphic inshore). Meanwhile, the deep basins tend to have clays and silts that are washed away from the bank areas and deposited on top of the deeper till and rock layers. The experimental site, NS1, is located at 44°05N, 61°10W on the Scotian shelf (see Fig. 11) approximately 45 nautical miles from the shelf edge. To avoid interfering sources of noise, it is away from the seismic exploration presently concentrated along the shelf edge and from the oil and gas production platforms (the nearest is Alma at 43°35.7N 60°40.9W).

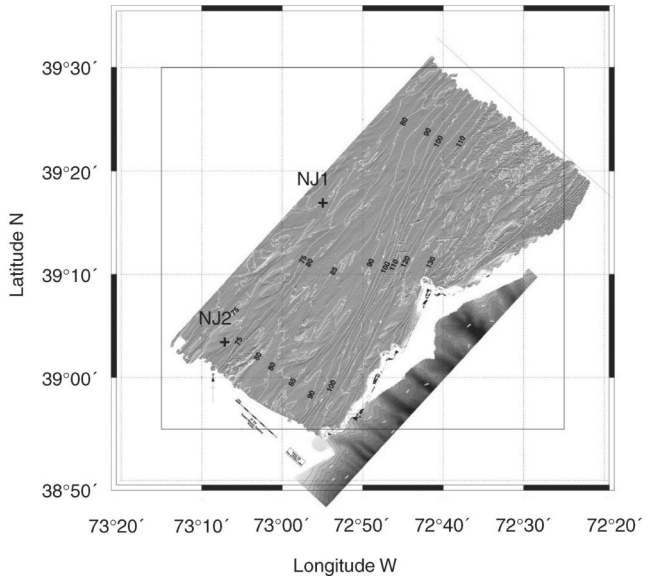


FIG. 10. Map showing the Strataform East area and experimental site NJ1.

To characterize the surficial geology at NS1, and examine the temporal variation of the surface roughness, a series of sidescan and sub-bottom profiler survey lines was collected 6 months prior to the scattering trial and again during the trial. Grab samples were also taken during the trial. The surficial sediment at NS1 is a well-sorted medium sand, 0.25 to 0.5 mm (ϕ=1.0 to 2.0). There is a near-surface reflector at a depth of 0 to 3 m beneath the seabed that marks the top of a layer of sand dominated by glacial deposits. The sand layer is approximately 75 m thick and is underlain by tertiary aged sedimentary rock. The WBS was situated in one of an extensive series of quasiparallel ribbon-like bedforms whose long axes are oriented north–northeast to south–southwest. The ribbons are slight depressions in the seabed caused by erosion due to the prevailing currents, and are typically hundreds of meters long, approximately 5 m wide, and 10 to 30 m apart from one another. Finer-grained material is removed from these depressions, leaving behind coarser sand. Low-amplitude ripples are sometimes present within these depressions, indicating that the area has active sediment transport, but none was present at the time of the scattering measurements. The ribbons were the only surface feature within the range of the WBS experiments. To determine the depth of the ribbons relative to the surrounding seabed, a swath bathymetry survey was conducted 1 year after the trial, using a 455-kHz Reson 8125. The ribbons were found to be in the same location as the previous two sea trials, suggesting that it would be appropriate to use the swath data despite the intervening period. The ribbons are typically 0.3 to 0.5 m deep. By examining the bathymetry perpendicular to and parallel to the ribbon axes, one obtains a standard deviation in the local grazing angle of 2.7° and 1.9°, respectively. Since the ribbon pattern is not deterministic, averaging across several azimuths will tend to smooth out range-dependent and compositional effects that may be manifest on a particular azimuth.

With the exception of the ribbons, the seabed is flat in the experimental area and there are no overall trends of any significance. Bottom loss measurements at 630, 1000, and 3150 Hz were combined with normal incidence seismic reflection data taken at the site to develop a two-layer model[22] for the seabed. A manual best fit to the data was obtained using a surficial layer 2.3 thick with a sound speed of 1647 ms^{-1} overlying a half-space with a sound speed of 1700 ms^{-1}.

V. THE EXPERIMENTAL GEOMETRY AND METHODOLOGY

A. *In situ* calibration

Due to space constraints it was not possible to calibrate the PATS head while it was mounted on the WBS space frame at the calibration barge. Therefore, in order to most closely replicate the experimental geometry, a subset of the calibration measurements were made prior to performing the

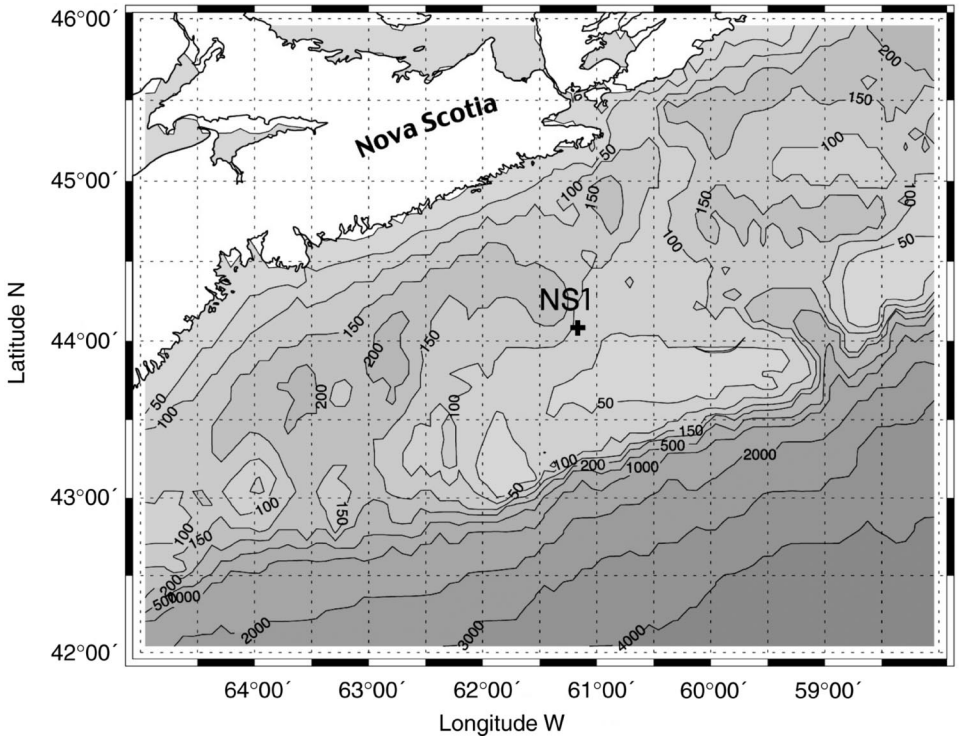


FIG. 11. Map showing the Scotian Shelf area and experimental site NS1.

scattering experiments. Figure 12 contains a sketch of the experimental geometry used to perform the *in situ* calibration. The WBS was deployed in bottom-mounted mode (recall Fig. 2) and DRDC Atlantic's Underwater Acoustic Target[23] (UAT)—an eight-hydrophone vertical line array (VLA)—was deployed approximately 160 m away from the WBS. The VLA hydrophones were spaced at 4.2-m intervals so that it could sample a substantial portion of the parametric array's vertical beam cross section. Both systems were localized in the horizontal plane to an accuracy of ± 3 m using the technique described in Ref. 24. Depth sensors on both systems were used to determine the vertical position of source and receivers. The PATS head was steered at a vertical angle of $+9°$ so that the beam axis was coincident with the hydrophone nearest the center of the UAT array. Then, a series of 50 pulses, 0.5 ms long was transmitted by the WBS at 4 kHz and the received level was recorded on all UAT hydrophones. To ensure that the VLA was stable during the measurements, the 50 pings were coherently averaged. For a Gaussian noise background one would expect a gain of $10\log(50)=17$ dB. If the VLA was moving relative to the transmitter, the coherent gain would be reduced. The measured coherent gain throughout the calibration was 17 ± 1 dB, indicating that the VLA was not moving relative to the parametric array. The PATS head was rotated 1.5°–2° in azimuth, and this sequence was repeated. Measurements were taken at

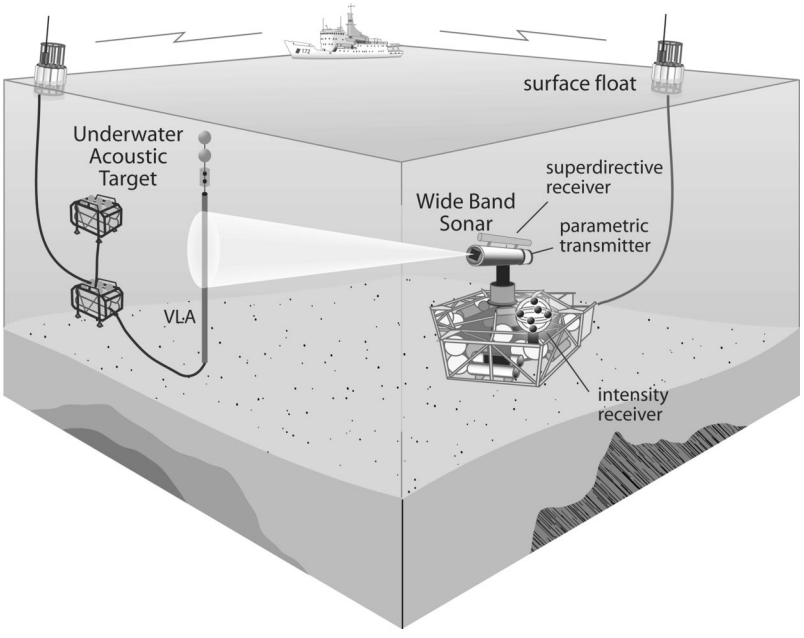


FIG. 12. Experimental geometry used to perform *in situ* calibration of parametric array.

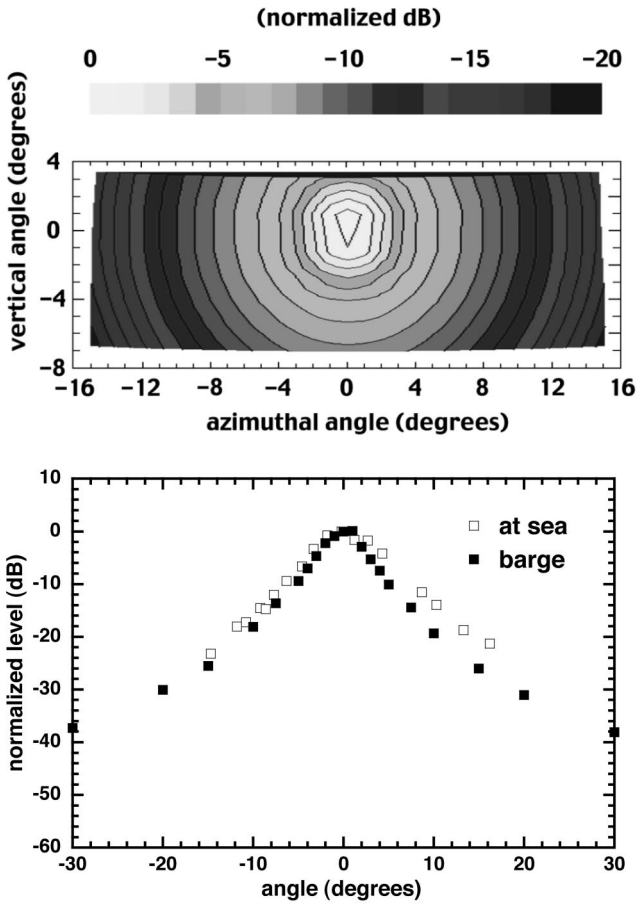


FIG. 13. *In situ* beam calibration results (a) Contour plot of azimuthal and vertical beam pattern obtained at sea. (b) Cross section comparing *in situ* beam pattern measurement to measurements obtained at DRDC Atlantic's calibration barge.

azimuths from −15° to +15° relative to the beam axis. This technique enabled an estimate of the vertical beam pattern using the aperture of the eight-hydrophone array and the horizontal beam pattern using the azimuthal scan. Figure 13(a) shows a contour plot of the *in situ* calibration data. Figure 13(b) shows the cross section of the beam pattern obtained on the center hydrophone during the azimuthal scan. The beam pattern measured at the calibration barge at 4 kHz is replotted from Fig. 9. The agreement between the two calibrations is reasonably good, but the *in situ* calibration suggests a two-sided half-power beamwidth of approximately 7°—compared to 4.5° measured at the calibration barge. Since the calibration at the barge was conducted in a very well-controlled environment, one would think the *in situ* calibration is suspect. The most obvious cause of error would be relative motion between source and receive array during the *in situ* calibration. This could occur, for example, if ocean currents caused the vertical line array to sway; however, the coherent average of 50 pings taken at a single azimuthal angle agreed to within 1 dB of the result obtained by computing the rms average for the same angle. If there was relative motion between the source and the receiver, one would not be able to coherently average the pings. Additionally, if there was relative motion between the WBS and the VLA during the course of the calibration, one would expect some smearing of the beam pattern; but, neither the cross section nor the surface plot in Fig. 13 show significant evidence of this. The *in situ* calibration provides SL_d at only a single range value since the separation between source and receiver could not be altered. The value obtained at 160-m range was approximately 188 dB re: 1 μPa@1 m. This source level is consistent with the asymtoptic nature of the data plotted in Fig. 6.

B. The scattering measurements

Figure 14 contains a sketch of the experimental geometry used to measure backscatter from the seabed. The WBS was deployed in bottom-mounted mode (recall Fig. 2) and the sonar was initially pointed at a grazing angle of 15°. A series of 50 pulses, 2–5 ms in duration, was transmitted by the WBS at 4 and 8 kHz, and acoustic backscatter from the

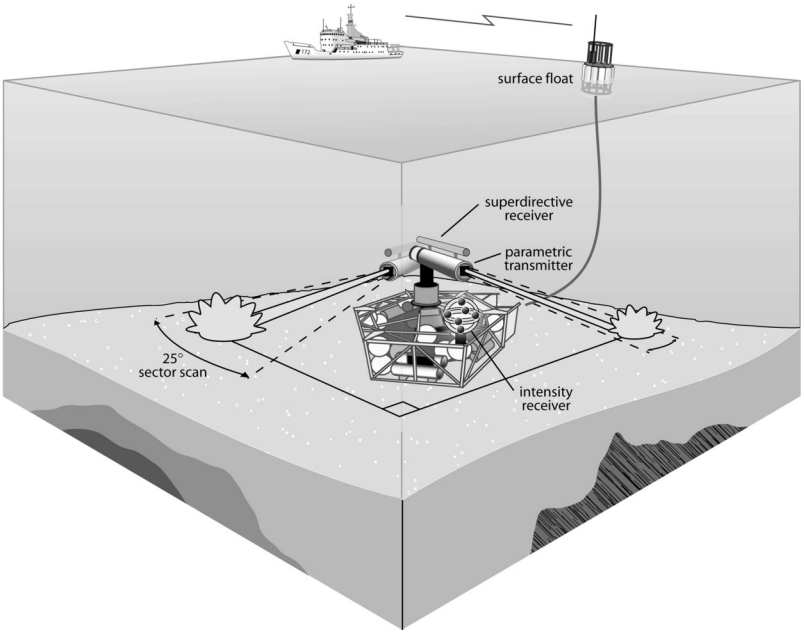


FIG. 14. Experimental geometry used to measure acoustic backscatter from the seabed.

seabed was recorded on SIREM. The parametric array transmitter was rotated in azimuth and the sequence was repeated. The backscattered intensity was measured at several azimuths, the results averaged, and converted to backscattering strength (BSS). The grazing angle was reduced and the procedure was repeated. Backscatter data were collected with the sonar pointing at grazing angles from 5° to 15°, resulting in coverage from about 3° to 15° grazing.

In addition to measuring the BSS as a function of grazing angle, a sector scan was made at 10° grazing angle to examine the feasibility of using the WBS to measure the azimuthal variability of acoustic scatter. Once again, a series of 50 pulses, 2 ms in duration, was transmitted by the WBS at 4 and 8 kHz, and the returns were recorded on SIREM. The parametric array transmitter was rotated approximately 1.5° to 2° in azimuth and the sequence was repeated. This sector scan covered approximately $\Delta\theta = 25°$. Time constraints did not permit a 360° sector scan; as a compromise, the parametric array was rotated 90° in azimuth relative to the center of the sector scan, and the experiment was repeated for two azimuthal angles separated by 3°. This procedure was adopted to allow an examination of the small-scale transition of backscatter across azimuth, while at the same time providing an opportunity to observe any drastic variations that might only show up with a substantial change in azimuth. The transmitter was rotated approximately 1.5° to 2° to ensure there was significant overlap of the area ensonified for adjacent measurements in the sector scan. As will be seen later in the paper, the overlap ensures a smooth transition for the visual (waterfall) display of the data. These measurements, combined with those made at 90° relative azimuth, provide between five and eight independent patches of seabed for the BSS calculation depending on whether the barge or the *in situ* calibration is correct.

In the following section, estimates of the grazing angle dependence of the BSS at 4 and 8 kHz obtained from the experiment are presented along with a sample of the azimuthal variability of the scattered intensity for the Strataform East site.

VI. RESULTS AND DISCUSSION

A. The backscattering strength measurements

The solid line in Fig. 15 shows a sample of the backscattered energy received on a single SIREM hydrophone plotted as a function of time. The data are at a slant angle of 10° and 8-kHz difference frequency. The initial transmit pulse (corresponding to $T=0$) and the main return centered at about 20 ms are labeled in the figure. As can be seen in Fig. 15, there is substantial SNR from about 15 to 30 ms. These times-of-flight map to grazing angles of $15° \geqslant \phi \geqslant 7°$ using the formula

$$\phi = \sin^{-1}\frac{h}{c_w T/2}, \tag{1}$$

where h is the sonar height above the seabed, c_w is the compressional wave speed in water at the depth of the WBS, and $c_w T/2$ is the slant range to the seabed corresponding to time T after transmission. [Note that Eq. (1) assumes a flat seabed which, within the limits discussed in Sec. IV, is consistent with the topography at the experimental sites.] Thus, the BSS can be computed for a reasonably wide range of grazing angles for a single slant angle of the sonar. This allows some overlap in angle for each of the BSS estimates obtained, which increases confidence in the results.

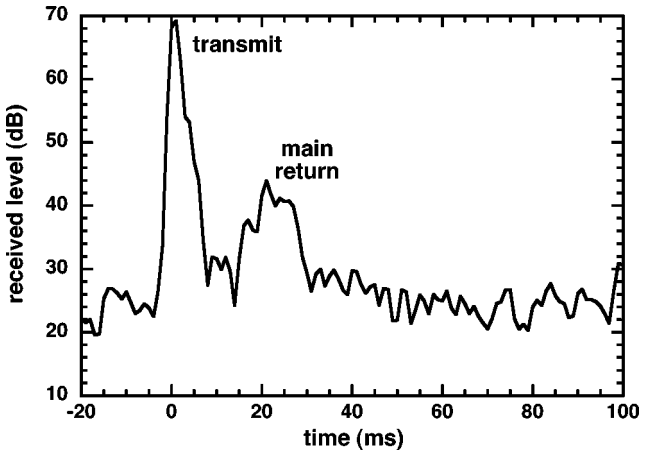


FIG. 15. A sample of the backscattered energy received on a single SIREM hydrophone plotted as a function of time. Time $T=0$ corresponds to the time of transmission of the pulse.

There is a risk that the received signal could be contaminated by scatter or reflections from the space frame as the wave propagates from the scattering patch back to the receiver. (This is not an issue for the outgoing parametric pulse because the outgoing pulse is too narrow to interact with the space frame.) The dimensions of the frame and the height of the sonar above the framework dictate that an angle of 30° below the horizontal is the shallowest angle from which returns from the ensonified patch of seabed could scatter or reflect from the frame and impinge on the receiver. These interfering returns would be approximately time coincident with the direct arrival because the path lengths would be almost identical. To address this concern, the single-hydrophone data were compared to the same returns processed using a fifth-order superdirective (SD) array.[25] The beam pattern of the fifth-order SD array is such that, relative to the on-axis response, the received sensitivity is -6 dB at 30° off-axis, and -10 dB at 35° off-axis. If significant energy was scattered/reflected by the frame, then it would show up in the single-hydrophone data but would be rejected by the superdirective processing. (If this occurs, the two data sets would not have the same received level on-axis, since the single-hydrophone results would include additional scatter paths, whereas the SD processing would reject them.) Figure 16(a) compares 4-kHz time series data from a single hydrophone (solid line) with the fifth-order SD results (dash line). Both curves represent the average across several azimuths. The data contained within the dashed box have been replotted in Fig. 16(b) in terms of equivalent[26] grazing angle ϕ using Eq. (1). Above about 20° grazing, the beam pattern of the SD array begins to reject the steeper paths, resulting in the lower level of the dash curve. However, near the center of the beam (say, from 15°–7°), the results measured on the SD array are within about 1.5 dB of the levels measured on the single hydrophone. This places an upper bound on con-

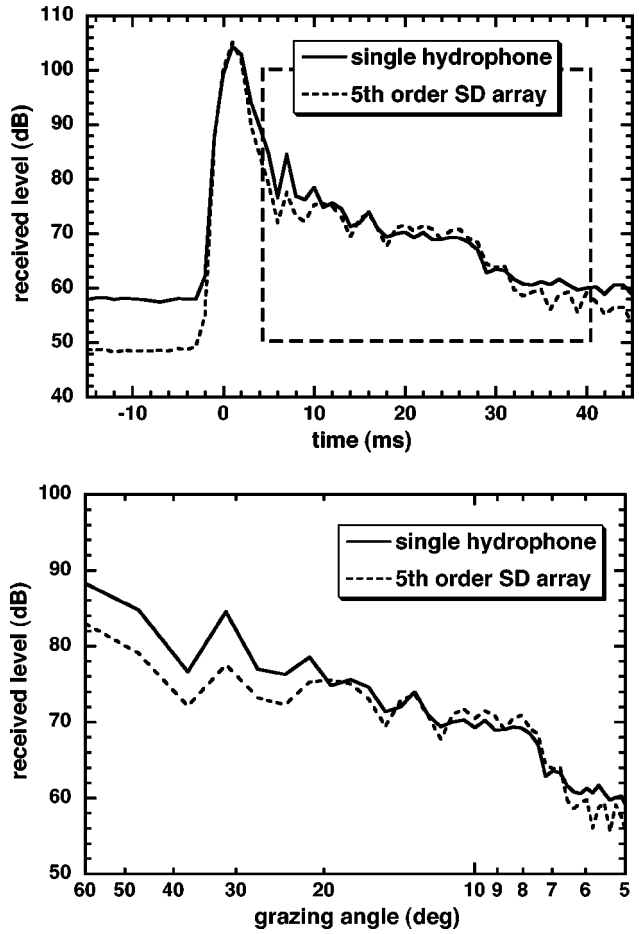


FIG. 16. (a) Comparison of 4-kHz time series data collected on a single SIREM hydrophone (solid line) with data processed superdirectively using all six hydrophones in the SIREM array (dashed line). (b) Replot of the data contained within the dashed box of (a) in terms of grazing angle. Note that a logarithmic horizontal axis has been used to enhance the display of shallow angles.

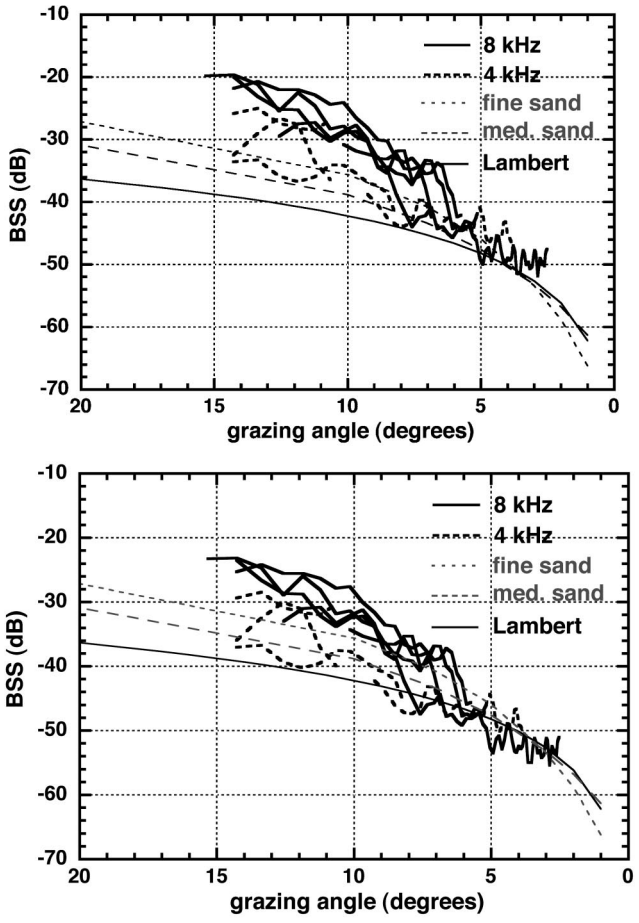


FIG. 17. (a) Backscattering strength (BSS) vs grazing angle for the Scotian Shelf site computed using barge calibration data. Results are given at 4 kHz (thick dashed lines) and 8 kHz (thick solid lines). The data are compared to model estimates of BSS from fine sand (thin dotted line) and medium sand (thin dashed line). Lambert's law with $\mu=-27$ dB is plotted for reference (thin solid line). (b) BSS vs grazing angle for the Scotian Shelf site computed using *in situ* calibration data. Results are given at 4 kHz (thick dashed lines) and 8 kHz (thick solid lines). The data are compared to model estimates of BSS from fine sand (thin dotted line) and medium sand (thin dashed line). Lambert's law with $\mu=-27$ dB is plotted for reference (thin solid line).

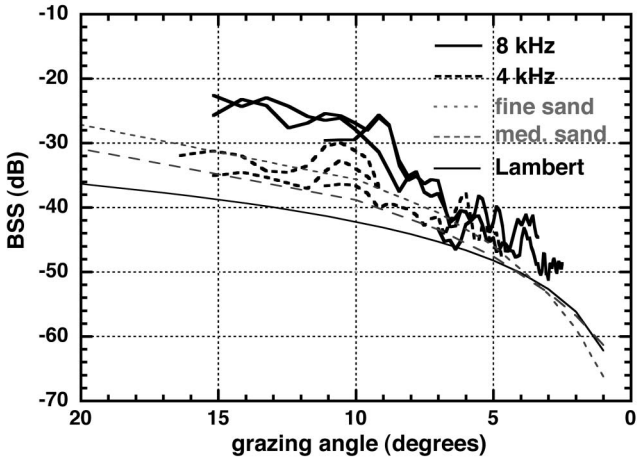
tamination from the space frame of about 1.5 dB. The separation of the two curves below 7° occurs because the single hydrophone is approaching the ambient noise floor, whereas the noise floor for the SD array is about 10 dB lower [recall time $T<0$ in Fig. 16(a)].

The energy time series was converted to BSS using the expression

$$\mathrm{BSS}=\mathrm{RL}-10\log(A)-\mathrm{SL}[f_d,R]+20\log(c_wT/2)-10\log(\tau), \quad (2)$$

where RL is the received intensity, A is the area ensonified by a pulse of duration τ and beamwidth Ω (measured in steradians), SL_d is the source intensity at difference-frequency f_d and range R, $20\log(c_wT/2)$ accounts for spherical spreading from the scattering patch to the receiver, and the term $10\log(\tau)$ is required to convert from units of energy to power. Values for SL_d and the beamwidth were obtained from the calibration data and models summarized in Figs. 7–9 and 13.

The backscattered energy was measured with the acoustic axis centered at slant angles of 5°, 7.5°, 10°, and 15°. To prevent discrete features from biasing the results, measurements were made at several azimuthal angles and the rms average was computed. The BSS was computed using Eq. (2). Equation (2) can also assist in providing a bound on the error in the BSS estimate. The source level and receive sensitivity were calibrated to within ±1 dB. The array was within a few meters of the seabed, where the sound speed was very stable during the course of the measurements. Therefore, errors related to the spreading term are minimal. [Errors caused by scatter from the frame will not appear in Eq. (2) but are less than 1.5 dB as discussed previously.] The largest source of uncertainty in the BSS occurs in the estimate of the ensonified area. The discrepancy in the beam pattern measurements made at the calibration barge and *in situ* during the experiment result in an uncertainty of approximately 3 dB. Figures 17 and 18 show the BSS at 4 and 8 kHz for the Scotian Shelf and Strataform East sites, respectively. The data are plotted as a function of grazing angle employing the barge calibration [Figs. 17–18(a)] and the *in situ* calibration [Figs. 17–18(b)]. The envelope of the data computed using both calibration estimates is shown in Figs.



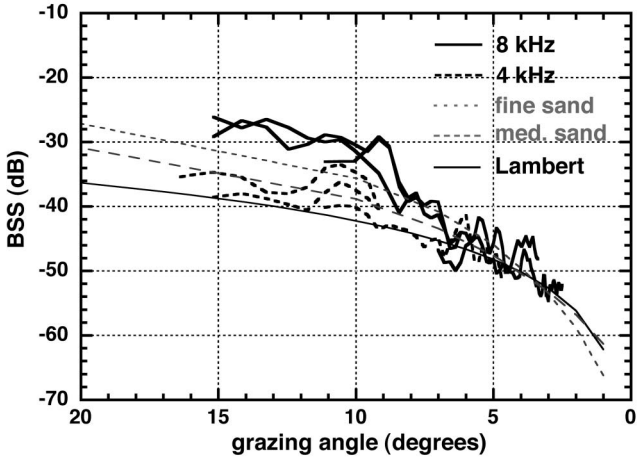


FIG. 18. (a) BSS vs grazing angle for the Strataform East site computed using barge calibration data. Results are given at 4 kHz (thick dashed lines) and 8 kHz (thick solid lines). The data are compared to model estimates of BSS from fine sand (thin dotted line) and medium sand (thin dashed line). Lambert's law with $\mu=-27$ dB is plotted for reference (thin solid line). (b) BSS vs grazing angle for the Strataform East site computed using *in situ* calibration data. Results are given at 4 kHz (thick dashed lines) and 8 kHz (thick solid lines). The data are compared to model estimates of BSS from fine sand (thin dotted line) and medium sand (thin dashed line). Lambert's law with $\mu=-27$ dB is plotted for reference (thin solid line).

19 (Scotian Shelf) and 20 (Strataform East). This allows one to obtain a useful uncertainty bound on the BSS when employing them in sonar performance modeling. Below about 8° grazing, the data at both sites exhibit little, if any, frequency dependence in the scattering strength. At both sites, the BSS at 4 kHz is about 6–10 dB lower than at 8 kHz, from 15°–10°; however, the levels of the 4- and 8-kHz data converge at shallower angles. The fact that the two sites exhibit similar frequency and grazing angle dependences in the BSS does not come as a surprise, given the similarities in sediment type and surficial geology. Both sites have a pervasive layer of sand. Although the sand overlays a more complicated geological structure, it is the sand layer that dominates the acoustic response.

Three model curves are also shown in Figs. 17–20 for comparison. First, Lambert's law, given by

$$\mathrm{LL}=\sigma+10\log[\sin^2(\phi)], \tag{3}$$

with a Lambert coefficient of $\sigma=-27$ dB, is shown as the

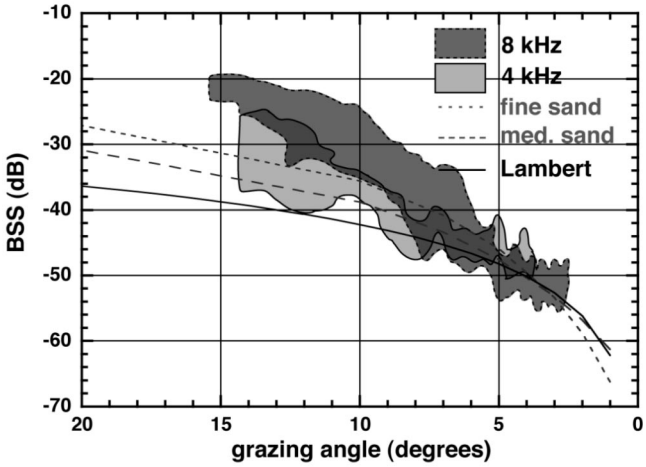


FIG. 19. Envelope of the BSS vs grazing angle data for the Scotian Shelf site at 4 and 8 kHz. The dark patches in the figure correspond to regions where the 4- and 8-kHz data overlap. The model estimates from Fig. 17 have been replotted in this figure.

thin solid line. Although, the angular dependence given by Lambert's law does not match the data, the levels are similar below 10° grazing. The remaining two curves were obtained from tabulated BSS values using the APL-UW high-frequency model[27] at a frequency of 10 kHz—the lowest frequency for which model results were given. The thin dotted line in the figure corresponds to the BSS tabulated for fine sand, and the thin dashed line corresponds to the BSS tabulated for medium grain sand. These curves represent a range of $\phi=1.5$ to 3.0, which is consistent with the grab samples taken in the area (recall Sec. IV of this paper). These curves should not be construed as an attempt to model the experimental data. It was not possible to measure several of the geophysical inputs required by the model. Furthermore, the model has only been tested down to 10 kHz. However, the agreement in slope and level between the data and the model supports the validity of the measurements.

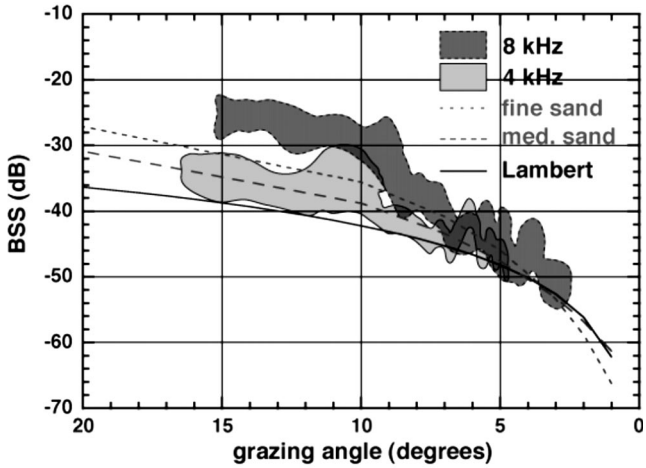


FIG. 20. Envelope of the BSS vs grazing angle data for the Strataform East site at 4 and 8 kHz. The dark patches in the figure correspond to regions where the 4- and 8-kHz data overlap. The model estimates from Fig. 18 have been replotted in this figure.

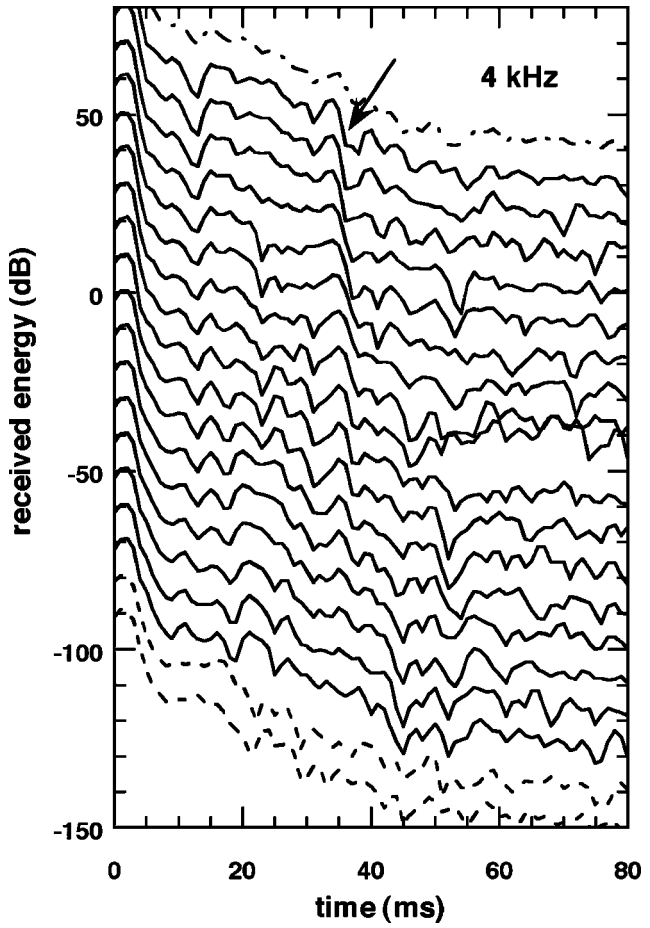


FIG. 21. Waterfall display of azimuthal dependence of backscattered energy vs time for the Strataform East site taken at 4 kHz; there is a 10-dB offset between each curve in the waterfall. The topmost (dashed) trace in the figure is the average for the entire sector scan. The two dashed lines at the bottom of the figure are measurements taken at azimuths 90° and 93° relative to the center of the sector scan.

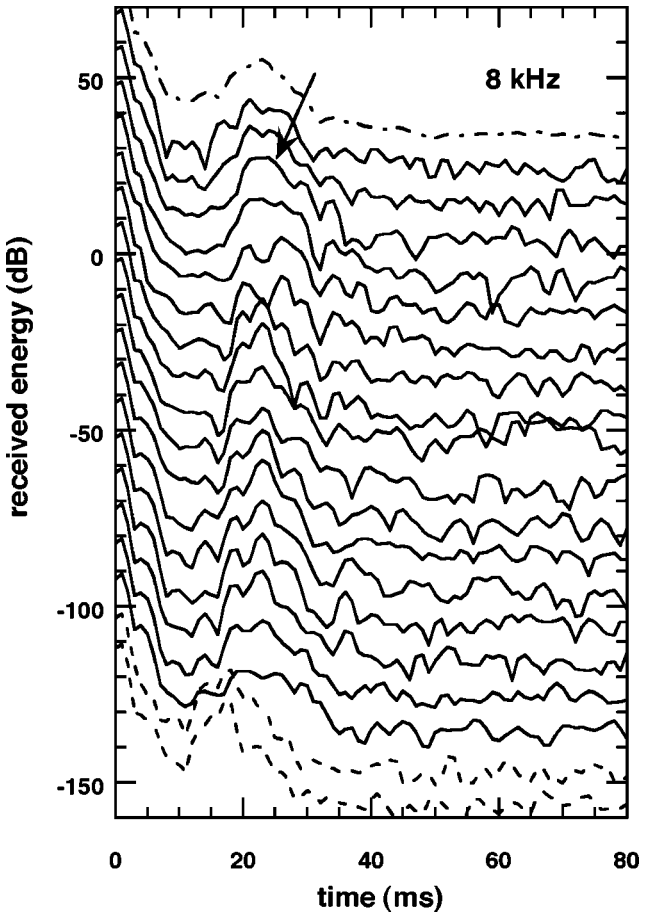


FIG. 22. Waterfall display of azimuthal dependence of backscattered energy vs time for the Strataform East site taken at 8 kHz; there is a 10-dB offset between each curve in the waterfall. The topmost (dashed) trace in the figure is the average for the entire sector scan. The two dashed lines at the bottom of the figure are measurements taken at azimuths 90° and 93° relative to the center of the sector scan.

B. The azimuthal variability

The results of the sector scan made at the Strataform East site are shown in the waterfall displays of Figs. 21 (4 kHz) and 22 (8 kHz). These data are backscattered energy vs time as measured on a single hydrophone in SIREM. The top (dashed) trace is the average of the 17 angles used in the sector scan. The 17 solid lines below it correspond to the data at each azimuthal angle. Each trace in the waterfall represents the coherent average of 50 pulses at a single azimuth. These data were taken with the parametric array pointed at 10° grazing. This corresponds to a two-way travel time of approximately 20 ms at the center of the beam. The two dashed lines at the bottom correspond to measurements at azimuths of 90° and 93° relative to the center of the sector scan. Unfortunately, experimental constraints required that for these measurements the sonar be pointed at 13° grazing (rather than 10° used during the rest of the sector scan). This corresponds to a two-way travel time of 15 ms to the center of the beam.

The waterfall display shows rich structure in the azimuthal dependence of the data. For example, the arrow in the 8-kHz data points to a single peak splitting in two and then consolidating back into a single peak. This occurs a couple of times as one passes through azimuth. Similar features occur in the 4-kHz data. Additionally at 4 kHz, a sharp ridge (denoted by the arrow) can be seen in the upper traces that disappears near the bottom of the figure. At this time it is unclear whether these features result from interface structure or shallow sub-bottom layering. Although this sector scan is limited to a relatively small range of angles, it does demonstrate the potential of the WBS to examine the azimuthal variability of seabed backscatter. Future experiments will address this in greater detail.

VII. CONCLUSIONS

The wideband sonar is an effective tool for examining seabed acoustics. It permits a wide range of experimental geometries, minimizes the risk of acoustic interference from the ship, and prevents ship motion from compromising array stability. Its narrow beamwidth makes the system particularly well suited to measure the variability of acoustic scattering from the seabed in addition to making direct (rather than inferred) measurements of the backscattering strength of the seabed.

In this paper, the azimuthal variability of the scattered intensity and the grazing angle dependence of the seabed

backscattering strength are reported for frequencies of 4 and 8 kHz. The measurements were made at two shallow-water sites: one on the New Jersey continental shelf (Strataform East) and one on the Scotian Shelf. The backscattering strength is similar at both sites. This is not surprising given the similarities in sediment type and surficial geology of the two sites. Below about 8°, the BSS appears to be independent of frequency within the statistical accuracy of the data. The measurements show reasonable agreement with model estimates of backscatter from sandy sediments. The limited results obtained on the azimuthal dependence of backscatter indicate that there is substantial azimuthal variability of seabed backscatter, and suggest that the wideband sonar is well suited to examining this in detail.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of Mr. Paul Shouldice, Mr. Roger Arsenault, and Mr. Victor Young of DRDC Atlantic in the collection and analysis of the scattering data used in this paper, Dr. Anna Crawford of DRDC Atlantic for her help in examining seabed roughness, Mr. Donald Glencross of DRDC Atlantic for his assistance with illustrations, and Professor Victor Humphrey of the University of Southampton for his insight into parametric array performance.

[1] D. D. Ellis, J. R. Preston, R. Hollett, and J. Sellschopp, ''Analysis of Towed Array Reverberation Data From 160 to 4000 Hz During Operation Rapid Response 97,'' SACLANT Centre Report, SACLANTCEN/SR-280 (2000).

[2] K. L. Williams, D. R. Jackson, E. I. Thorsos, D. Tang, and K. B. Briggs, ''Acoustic backscattering experiments in a well characterized sand sediment: Data/model comparisons using sediment fluid and Biot models,'' IEEE J. Ocean. Eng. **27**, 376–387 (2002).

[3] A. V. Bunchuk and Yu. Yu. Zhitkovskii, ''Scattering by the ocean bottom in shallow-water regions,'' Sov. Phys. Acoust. **26**(5), 363–370 (1980).

[4] C. W. Holland, ''Coupled scattering and reflection measurements in shallow water,'' IEEE J. Ocean. Eng. **27**, 454–470 (2002).

[5] C. W. Holland, R. Hollett, and L. Troiano, ''Measurement technique for bottom scattering in shallow water,'' J. Acoust. Soc. Am. **108**, 997–1011 (2000).

[6] R. J. Soukup and R. F. Gragg, ''Backscatter from a limestone seafloor at 2–3.5 kHz: Measurements and modeling,'' J. Acoust. Soc. Am. **113**, 2501–2514 (2003).

[7] G. Frisk, R. Gauss, C. Holland, N. Makris, and A. Malekhanov (Guest Editors), ''The interaction of low-to-mid frequency sound with the ocean bottom,'' Special Issue, J. Oceanic Eng. (to be published).

[8] C. M. Day and T. Yamamoto, ''Low grazing angle bistatic sea floor scattering on the Florida Atlantic Coastal Shelf,'' J. Acoust. Soc. Am. **106**, 1744–1754 (1999).

[9] P. C. Hines, W. Cary Risley, and M. P. O'Connor, ''A Wide-Band Sonar for Underwater Acoustics Measurements in Shallow Water,'' Proceedings: Oceans'98, **3**, 1558–1562 (1998).

[10] P. C. Hines, ''Quantifying seabed properties in shelf waters using a parametric sonar,'' Meas. Sci. Technol. **10**, 1127–1134 (1999).

[11] P. J. Westervelt, ''Parametric acoustic array,'' J. Acoust. Soc. Am. **35**, 535–537 (1963).

[12] M. B. Moffett and R. H. Mellen, ''Model for parametric acoustic sources,'' J. Acoust. Soc. Am. **61**, 325–337 (1977).

[13] P. C. Hines, A. L. Rosenfeld, B. H. Maranda, and D. L. Hutt, ''Evaluation of the Endfire Response of a Superdirective Line Array in Simulated Ambient Noise Environments,'' Proceedings: Oceans'00, Providence, RI, 11–14 Sept., 2000, pp. 1489–1494.

[14] D. L. Hutt, P. C. Hines, and V. W. Young, ''Impact of Measurement Error and Noise on Underwater Sound Intensity,'' Proceedings: 7th International Conference on Sound and Vibration, 4–7 July, 2000, Garmisch, Germany.

[15] V. F. Humphrey and C. H. Hsu, ''Nonlinearity of cylindrical hydrophones used for the measurement of parametric arrays,'' in Proceedings of the Institute of Acoustics Conference on Transducers for Sonar Applications, University of Birmingham, December 1980.

[16] R. H. Mellen and M. B. Moffett, ''A numerical method for calculating the near field of a parametric acoustic source,'' J. Acoust. Soc. Am. **63**, 1622–1624 (1978).

[17] J. A. Goff, D. J. P. Swift, C. S. Duncan, L. A. Mayer, and J. Hughes-Clarke, ''High-resolution swath sonar investigation of sand ridge, dune, and ribbon morphology in the offshore environment of the New Jersey Margin,'' Mar. Geol. **161**, 307–337 (1999).

[18] J. C. Osler, ''Geoacoustic characterization of Boundary 2001 experimental locations,'' Proceedings: Acoustic Interaction with the Seabed, Conference held at DREA, Canada 2–3 October, 2001, DREA TM-2001-185.

[19] C. S. Duncan, J. A. Goff, J. A. Austin, Jr., and C. S. Fulthorpe, ''Tracking the last sea-level cycle: Seafloor morphology and shallow stratigraphy of the latest Quaternary New Jersey middle continental shelf,'' Mar. Geol. **170**, 395–421 (2000).

[20] L. A. Mayer, B. J. Kraft, P. Simpkin, P. Lavoie, E. Jabs, and E. Lynskey, ''*In situ* determination of the variability of seafloor acoustic properties: An example from the ONR geoclutter area,'' in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer Academic, Dordrecht, Netherlands, 2002), pp. 115–122.

[21] B. J. Kraft, L. A. Mayer, P. Simpkin, P. Lavoie, E. Jabs, E. Lynskey, and J. A. Goff, ''Calculation of *in-situ* acoustic wave properties in marine sediments,'' in Ref. 20, pp. 123–130.

[22] J. C. Osler and D. D. Ellis, ''Geological clutter on Sable Island bank,'' IEEE J. Ocean. Eng. (submitted).

[23] P. C. Hines, J. C. Osler, and D. L. Hutt, The Environmental Acoustics group's seagoing measurement systems, DREA TM 2001-173, 2001.

[24] J. C. Osler and J. Beers, ''Real-time localization of multiple acoustic transponders using a towed interrogation transducer,'' Proceedings: Oceans 2000, Providence, RI, 11–14 September (2000).

[25] V. F. Humphrey, P. C. Hines, and V. Young, ''Experimental Analysis Of The Performance Of A Superdirective Line Array,'' Proceedings: The Seventh European Conference on Underwater Acoustics, ECUA 2004, Delft, The Netherlands, 5–8 July, 2004.

[26] It is important to distinguish between the ''true'' steep angle paths scattered from the seabed and shallow angle paths that are subsequently scattered from the frame, into steep angle paths. The former map to steep angles using Eq. (1), whereas the latter arrive at approximately the same time as the shallow angle paths coming directly from the ensonified patch and consequently map to shallow angles.

[27] *APL-UW High-frequency Ocean Environmental Acoustic Models Handbook* (Applied Physics Laboratory-University of Washington, Seattle, APL/UW/TR-9407, 1994).

Application of the rational function approximation to Wiener–Hopf calculations for fluid-loaded plates

M. L. Rumerman
Naval Surface Warfare Center, Carderock Division Ship Signatures Department (Code 7204), 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5700

(Received 19 October 2004; revised 10 March 2005; accepted 11 March 2005)

This paper discusses the Wiener–Hopf analysis for the solution of fluid-loaded plate problems when the planar acoustic impedance is represented by a rational function approximation. With such a formulation, the factorization of the integrand kernel can immediately be written once the system poles are determined. The approach is applied to several representative problems, and the approximate results are shown to be in agreement with those of exact solutions. [DOI: 10.1121/1.1899883]

PACS numbers: 43.30.Jx, 43.40.Rj [ANN] Pages: 3517–3523

I. INTRODUCTION

Estimation of the broadband response characteristics of connected finite plates using power flow methods requires knowledge of the power scattering coefficients at the junctions of adjacent plates and at exterior boundaries. When the plates are large enough, or the frequency high enough, to justify broadband estimates in the first place, the power coefficients may be estimated by considering adjacent plates to each be semi-infinite and their common edge to be of infinite length. If the adjacent plates are dissimilar, and the analysis is performed in wave number (Fourier transform) space, the Wiener–Hopf or some analogous method must be used. Its implementation requires the factorization of a kernel function into two parts, one that is analytic and nonzero in an upper-half plane (UHP) of the Fourier transform variable, and one that is analytic and nonzero in a lower-half plane (LHP). When the plates are fluid loaded, the factorization as a function of wave number cannot be effected by inspection, and available procedures require integrations to evaluate the factors for each value of the argument.[1–4] Therefore, if the problem solution involves integrals whose evaluation includes not only residue contributions, but branch cut integrals whose integrands contain the UHP or LHP kernel factors, the arithmetic can be complicated and subject to inaccuracies.

In Ref. 4, Abrahams proposed using a rational function (ratio of two polynomials) approximation (RFA) to the combined plate and fluid impedances. His objective was to develop an accurate approximation to the true factorization. With such a representation, the factorization can be done by inspection after finding the roots of a polynomial. To obtain this representation, Abrahams first developed an RFA of the acoustic impedance using a Padé approximant based on a Taylor series expansion about zero wave number (i.e., in the supersonic neighborhood). Therefore, it is expected that the approach will be effective only well above the coincidence frequency; indeed, the excellent comparison between the exact and approximate factorization presented by Abrahams was calculated for a frequency that is four times the coincidence frequency. He did not present an analysis of a representative fluid-loaded plate problem, and it is not clear how the approximate factorization is to be used.

More recently, DiPerna and Feit[5] used an alternative RFA formulation to calculate the response of a fluid-loaded plate to a line force. They determined the RFA for the acoustic impedance by enforcing a least-square fit to the values of the exact impedance at a large number of wave numbers along the real axis, in the supersonic region and in the subsonic region out to 125% of the bending wave number at the lowest frequency of interest. Using the poles and residues of the approximate representation, DiPerna and Feit calculated the normal velocity of a steel plate in water as a function of distance from a line force (for a range of frequencies), and the pressure on an infinite impedance surface as a function of distance from a line volume source. Because their results agreed very well with those of exact solutions, it was considered worthwhile to examine the suitability of their approximate model for Wiener–Hopf problems. The intent is not to develop an accurate approximation to the exact factors, but to use their model as an alternative representation of the plate–fluid system. The purpose of the present paper is to apply this approach to several benchmark problems, and to show that it is successful.

II. SYSTEM POLES AND KERNEL FACTORIZATION

Consider a plate with mass per unit area M and *in vacuo* bending wave number k_b at frequency $\omega/2\pi$ that is in contact with an acoustic fluid of mass density ρ and sound speed c. The system impedance Z, in wave number (Fourier transform) space, is given by Eq. (1a), in which $\overline{\alpha}$ and $\overline{\beta}$ are the wave numbers conjugate to the x- and y axes in the plane of the plate, $\varepsilon=1$ for one-sided fluid loading and $\varepsilon=2$ for two-sided loading, $k=\omega/c$, and $\exp(-i\omega t)$ time dependence is assumed.

$$Z(\overline{\alpha},\overline{\beta})=-i\omega M\left[1-\frac{(\overline{\alpha}^2+\overline{\beta}^2)^2}{k_b^4}\right]+\frac{\varepsilon\rho c}{\sqrt{1-\dfrac{(\overline{\alpha}^2+\overline{\beta}^2)}{k^2}}}. \tag{1a}$$

TABLE I. Coefficients for $S=10$ RFA of acoustic impedance.

q	Real (a_q)	Imag (a_q)	Real (b_q)	Imag (b_q)
0	0.999 995 670 0	−0.000 002 477 7	1.000 000 000 0	0.000 000 000 0
1	−4.051 579 800 0	1.541 887 800 0	−4.551 841 000 0	1.541 699 800 0
2	5.531 463 700 0	−5.497 959 800 0	7.434 874 900 0	−6.266 235 000 0
3	−2.200 833 500 0	7.159 251 600 0	−4.531 644 600 0	9.700 416 800 0
4	−1.164 789 700 0	−3.919 502 000 0	−0.531 757 920 0	−6.866 281 200 0
5	1.032 800 600 0	0.637 584 290 0	1.685 550 200 0	1.925 360 700 0
6	−0.121 357 980 0	0.078 586 175 0	−0.513 459 530 0	0.015 715 481 0
7	−0.023 498 180 0	0.000 015 160 2	0.009 188 944 7	−0.040 156 943 0
8	−0.002 144 837 8	0.000 069 946 7	−0.000 907 917 8	−0.010 351 777 0
9	−0.000 005 934 7	−0.000 000 105 7	−0.000 000 791 5	−0.000 168 345 0
10	⋯	⋯	0.000 000 002 4	−0.000 000 074 8

Upon normalizing the wave numbers with respect to the acoustic wave number k (e.g., $\alpha=\overline{\alpha}/k$), the impedance can be recast as in Eq. (1b), in which $\gamma^2=\alpha^2+\beta^2$, $\Omega=k^2/k_b^2$ is the ratio of the frequency to the bending coincidence frequency, and $\sigma\equiv\rho c/\omega M$. For steel plates in water, $\sigma\approx 0.13/\Omega$.

$$Z(\alpha,\beta)=-i\omega M\left[(1-\Omega^2\gamma^4)+\frac{i\varepsilon\sigma}{\sqrt{1-\gamma^2}}\right]. \tag{1b}$$

DiPerna and Feit approximated the acoustic impedance term by an RFA, as in Eq. (2), and determined the coefficients $\{a\}$, $\{b\}$ using a least-squares fit.

$$\frac{1}{\sqrt{1-\gamma^2}}\approx\frac{\Sigma_{q=0}^{S-1}a_q\gamma^{2q}}{1+\Sigma_{s=1}^{S}b_s\gamma^{2s}}=\frac{\overline{N}(\gamma)}{\overline{D}(\gamma)}. \tag{2}$$

Note that $\overline{N}(\gamma)/\overline{D}(\gamma)=O(1/\gamma^2)$ for $|\gamma|\gg 1$, in contrast to $1/\sqrt{1-\gamma^2}=O(1/\gamma)$. This means that the RFA Wiener–Hopf solution will require explicit consideration of the tangential derivative of the surface pressure (i.e., the tangential component of the acoustic particle velocity) at the junction or edge, whereas the exact formulation treats this implicitly and automatically. Because the RFA is an even function of γ, the approximate physical system maintains reciprocity.

The system impedance is approximated by Eqs. (3). The system poles are the $(2S+4)$ values of γ for which $\overline{W}(\gamma)=0$.

$$Z(\alpha,\beta)\approx -i\omega M\frac{\overline{W}(\gamma)}{\overline{D}(\gamma)}, \tag{3a}$$

$$\overline{W}(\gamma)=[(1-\Omega^2\gamma^4)\overline{D}(\gamma)+i\varepsilon\sigma\overline{N}(\gamma)]=0. \tag{3b}$$

Because the left-hand side of Eq. (3b) is an even function of γ, the poles may be given as $\pm\gamma_j$; $j=S+2$, where $+\gamma_j$ is defined as a UHP pole.

With one possible exception, the UHP poles are defined to have positive imaginary parts. The possible exception corresponds to the pole that represents the subsonic bending wave propagating in the positive-x direction, which will be designated γ_p. In the exact formulation, this pole is real and positive when the plate bending modulus is real. Because the DiPerna–Feit RFA representation is basically an empirical fit, it introduces a small real part to the acoustic impedance for subsonic real wave numbers; therefore, the pole γ_p will have a small imaginary part that can be positive or negative. Regardless of the sign of $\mathrm{Im}(\gamma_p)$, the contours that define the lower boundary of the UHP and the upper boundary of the LHP are drawn so that each runs beneath γ_p and above $-\gamma_p$, and so that the two half planes have a common overlapping band.

Below the coincidence frequency ($\Omega<1$), the pole γ_p is easily identified among the $2S+4$ solutions by normalizing the dimensional wave number with respect to k_b and choosing the one for which $\mathrm{Re}(\gamma_j\sqrt{\Omega})>+1$ and $|\mathrm{Im}(\gamma_j)|<\mathrm{Re}(\gamma_j)$. The calculations in this paper were performed with an $S=10$ RFA using values of $\{a\}$ and $\{b\}$ provided by DiPerna and listed in Table I. Table II compares the approximation of $\gamma_p\sqrt{\Omega}$ to values obtained from the exact formulation. The agreement of real parts is excellent, and the imaginary part of the approximation is very small.

The input admittance, $Y(\beta)$, for a normal line force $F\delta(x)\exp(ik\beta y)$ can be written as an integral over α using the system impedance in Eq. (3a). Its evaluation may be expressed as the sum of residues corresponding to the UHP or (LHP) poles. The specific contribution of the pole $\alpha_p=\sqrt{(\gamma_p^2-\beta^2)}$ is needed to determine the power scattering coefficients, and it is given by Eq. (4), in which the prime denotes a derivative with respect to γ and $d\gamma/d\alpha=\alpha/\gamma$

$$Y_p(\beta)=-\frac{k}{\omega M}\frac{\overline{D}(\gamma_p)}{\overline{W}'(\gamma)(d\gamma/d\alpha)|_{\gamma=\gamma_p}}. \tag{4}$$

As discussed in the Introduction, implementation of the Wiener–Hopf analysis requires that the kernel function be factored into UHP and LHP components. The kernel for the

TABLE II. Comparison of subsonic poles derived from RFA and exact formulation.

Ω	Exact $\gamma_p\sqrt{\Omega}$	RFA $\gamma_p\sqrt{\Omega}$
0.0100	0.120E+01	0.121E+01+0.974iE−03
0.0200	0.116E+01	0.116E+01+0.148iE−02
0.0398	0.112E+01	0.113E+01+0.134iE−02
0.0794	0.109E+01	0.110E+01+0.541iE−03
0.1585	0.107E+01	0.107E+01+0.424iE−04
0.3162	0.106E+01	0.106E+01−0.557iE−08
0.3981	0.106E+01	0.106E+01+0.754iE−08
0.5012	0.105E+01	0.105E+01−0.439iE−07
0.7943	0.106E+01	0.106E+01+0.978iE−05

problems in this paper is a polynomial, $\overline{W}(\gamma)$, and determination of its poles allows it to be written as the product of quadratic factors

$$\overline{W}(\gamma) = -\Omega^2 b_S \prod_{j=1}^{S+2} (\gamma^2 - \gamma_j^2). \tag{5a}$$

Because the actual variable of integration will be α, $\overline{W}$ is rewritten as a function of α

$$\overline{W}(\gamma) = -\Omega^2 b_S \prod_{j=1}^{S+2} [(\alpha^2+\beta^2) - \gamma_j^2)]$$

$$= -\Omega^2 b_S \prod_{j=1}^{S+2} (\alpha^2 - \alpha_j^2) = W(\alpha), \tag{5b}$$

where $\alpha_j^2 = \gamma_j^2 - \beta^2$ and β is a constant parameter. The UHP and LHP factors of $W(\alpha)$, respectively, $W^+(\alpha)$ and $W^-(\alpha)$, can be found by inspection. These are given by Eqs. (6a) and (6b), where $\{\alpha_j\}$ are the UHP poles corresponding to the UHP poles $\{\gamma_j\}$ as previously defined

$$W^+(\alpha) = \sqrt{-\Omega^2 b_S} \prod_{j=1}^{S+2} (\alpha + \alpha_j), \tag{6a}$$

$$W^-(\alpha) = \sqrt{-\Omega^2 b_S} \prod_{j=1}^{S+2} (\alpha - \alpha_j). \tag{6b}$$

It is convenient to define the factors so that $W^-(-\alpha) = W^+(+\alpha)$. Equations (6) automatically meet this condition when S is an even integer, as it is for the calculations in this paper. When S is odd, it can be satisfied by premultiplying the right-hand sides of Eqs. (6a) and (6b) by $+i$ and $-i$, respectively.

III. WIENER–HOPF PROBLEMS

The objective of this paper is to demonstrate the effectiveness of the DiPerna–Feit RFA model for water-loaded steel plates in Wiener–Hopf analyses. This will be done by comparing results for representative configurations with those of benchmark solutions. One class of problems considers a bending wave that freely propagates on one plate and is incident upon the junction with another plate. The incident wave satisfies the coupled fluid–plate equations for the first plate, but not for the second. The analysis involves deriving a secondary field scattered at the junction that, in combination with the incident wave, satisfies the equations for the second plate, and by itself satisfies the equations for the first plate. Procedures for fluid-plate problems are explained elsewhere,[1,2,5] and only major steps will be given here.

A. Two semi-infinite plates

Figure 1 shows two semi-infinite plates joined along the line $x=0$. The plates are in contact with an acoustic fluid, either on one side only ($\varepsilon=1$) or on both sides ($\varepsilon=2$), and a bending wave freely propagating on plate L ($x<0$) is incident with angle φ on the junction with plate R ($x>0$). The incident plate normal velocity is $V_0 \exp(ik\alpha_{pL}x)\exp(ik\beta y)$, where $\alpha_{pL} = \gamma_{pL}\cos\varphi$ and $\beta = \gamma_{pL}\sin\varphi$. The associated incident pressure field is

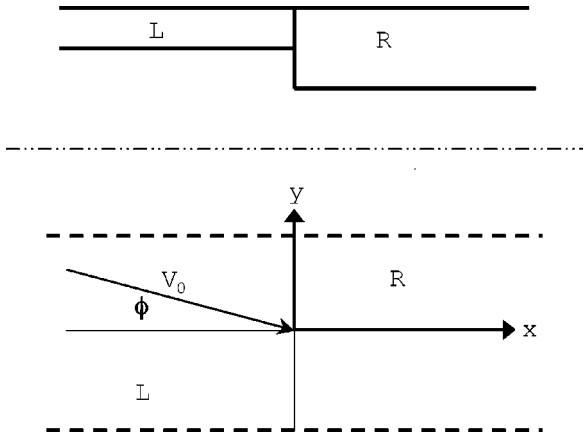


FIG. 1. Plate geometry.

$V_0\rho c\{[\overline{N}(\alpha_{pL})]/[\overline{D}(\alpha_{pL})]\}\exp(ik\alpha_{pL}x)\exp(ik\beta y)$, where $N(\alpha) = \overline{N}(\sqrt{\alpha^2+\beta^2})$ and $D(\alpha) = \overline{D}(\sqrt{\alpha^2+\beta^2})$. The Wiener–Hopf analysis follows that of Ref. 6, but with changes in notation, normalization of wave number by acoustic wave number rather than plate bending wave number, and use of the RFA for acoustic impedance. In what follows, the $\exp(ik\beta y)$ dependence is suppressed.

The field scattered from the edge is described in terms of the scattered pressure field, $P_s(x)$, which is given in terms of its as-yet unknown wave number spectrum J

$$P_s(x) = k\int_{-\infty}^{+\infty} J(\alpha)\exp(ik\alpha x)d\alpha. \tag{7a}$$

The associated scattered normal velocity is given by Eq. (7b)

$$V_s(x) = \frac{k}{\rho c}\int_{-\infty}^{+\infty} J(\alpha)\frac{D(\alpha)}{N(\alpha)}\exp(ik\alpha x)d\alpha. \tag{7b}$$

These fields must jointly satisfy the coupled fluid–plate equations for plate L, which produces Eq. (8a), in which the subscript L signifies plate L

$$\frac{-i\omega m_L k}{\rho c}\int_{-\infty}^{+\infty}\frac{W_L(\alpha)}{N(\alpha)}J(\alpha)\exp(ik\alpha x)d\alpha = 0; \quad x<0. \tag{8a}$$

Therefore

$$\frac{W_L(\alpha)}{N(\alpha)}J(\alpha) = \Lambda(\alpha), \tag{8b}$$

where $\Lambda(\alpha)$ is a function of α that is analytic in an LHP. The scattered fields combined with the incident waves must also satisfy the coupled equations for plate R, yielding Eq. (9a), in which the subscript R signifies plate R and the plus superscript on α_{pL}^+ emphasizes that it is a UHP pole

$$\frac{-i\omega m_R k}{\rho c}\int_{-\infty}^{+\infty}\left[\frac{V_0}{2\pi i}\frac{\rho c}{k}\frac{1}{(\alpha-\alpha_{pL}^+)}\frac{W_R(\alpha_{pL})}{D(\alpha_{pL})} + \frac{J(\alpha)W_R(\alpha)}{N(\alpha)}\right]\exp(ik\alpha x)d\alpha = 0; \quad x>0. \tag{9a}$$

Therefore

$$\frac{V_0}{2\pi i}\frac{\rho c}{k}\frac{1}{(\alpha-\alpha_{pL}^{+})}\frac{W_R(\alpha_{pL})}{D(\alpha_{pL})}+\frac{J(\alpha)W_R(\alpha)}{N(\alpha)}=U(\alpha), \quad (9b)$$

where $U(\alpha)$ is a function of α that is analytic in a UHP. The left-hand sides of Eqs. (8b) and (9b) may be factored into UHP and LHP components, rearranged, and combined to produce an equality of two expressions, one of which is analytic in a UHP and the other in a LHP, with a common strip of analyticity. This leads to a solution for $J(\alpha)$, in which the coefficients $\{B_n\}$ are constants whose values will be determined by the boundary conditions at the junction

$$J(\alpha)=\frac{V_0}{2\pi}\sum_{n=0}^{\max}B_n\frac{\alpha^n N(\alpha)}{W_L^{+}(\alpha)W_R^{-}(\alpha)}-\frac{V_0}{2\pi i}\frac{\rho c}{k}\frac{1}{(\alpha-\alpha_{pL}^{+})}\frac{W_L^{+}(\alpha_{pL})W_R^{-}(\alpha_{pL})}{D(\alpha_{pL})}\times\frac{N(\alpha)}{W_L^{+}(\alpha)W_R^{-}(\alpha)}. \quad (10a)$$

In this equation, max+1 is the allowable number of coefficients, $\{B_n\}$. The requirement $W_L^{+}(\alpha)W_R^{-}(\alpha)=W_L^{-}(-\alpha)W_R^{+}(-\alpha)$ ensures that the correct result is obtained as the plates become identical. The transform of the scattered normal velocity field is given by Eq. (10b)

$$\overline{V_s}(\alpha)=\frac{1}{\rho c}\frac{D(\alpha)}{N(\alpha)}J(\alpha). \quad (10b)$$

Four constants are required to satisfy the boundary conditions at the edges of the two plates at their junction, and this must equal the number of constants allowed through consideration of pressure requirements. From the definitions of $W(\alpha)$, $N(\alpha)$, and $D(\alpha)$ it can be seen that $[\alpha^n N(\alpha)]/[W_L^{+}(\alpha)W_R^{-}(\alpha)]\rightarrow\alpha^n/\alpha^6$; $|\alpha|\rightarrow\infty$. Continuity of pressure as a function of x requires that max be 4, but this leads to five undetermined constants. An additional requirement, discussed in the Appendix, is that the derivative of pressure with respect to x also be continuous. This reduces max to 3, and the number of constants to four.

The total plate velocity field is given by Eq. (11)

$$v(x)=V_0\exp(ik\alpha_{pL})+\frac{k}{\rho c}\int_{-\infty}^{+\infty}J(\alpha)\frac{D(\alpha)}{N(\alpha)}\exp(ik\alpha x)d\alpha. \quad (11)$$

The two plates will each be assumed to be clamped (i.e., zero normal velocity and zero slope) at their junction. It can be shown that, because $\overline{V_s}(\alpha)\rightarrow[\alpha^n D(\alpha)]/[W_L^{+}(\alpha)W_R^{-}(\alpha)]\rightarrow\alpha^n/\alpha^4$ as $|\alpha|\rightarrow\infty$, the coefficients B_2 and B_3 both vanish, and that if the coefficients B_0 and B_1 are chosen to satisfy the boundary conditions on plate L, they will automatically do so for plate R. These coefficients can be determined by using Eq. (11) to compute the velocity and slope of each plate as $x\rightarrow 0$ for that plate. The integrations can be performed using the theory of residues. For $x>0$ (plate R), the UHP poles are included and for $x<0$ (plate L), the LHP poles are used.

When the coefficients B_0 and B_1 have been determined, the amplitudes of the reflected and transmitted bending waves, $|V_{\text{ref}}|$ and $|V_{\text{tr}}|$, can be found from the coefficients of $\exp(-ik\alpha_{pL}x)$ and $\exp(+ik\alpha_{pR}x)$, respectively. The power associated with a propagating bending wave of amplitude $|V_p|$ is $|V_p^2|/[2\,\text{Re}(Y_p)]$; therefore, the power reflection coefficient is

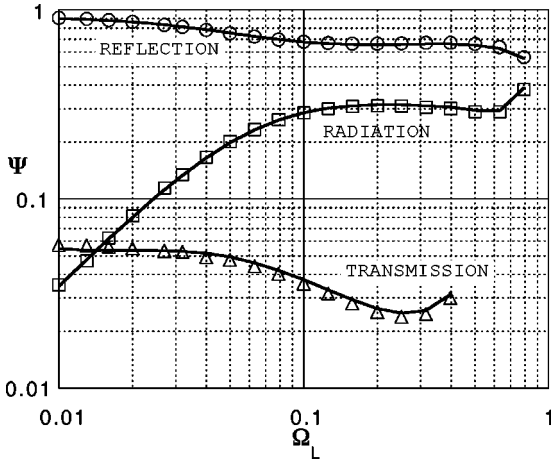


FIG. 2. Power scattering coefficients for two plates clamped along the junction (normal incidence). Solid lines: RFA; discrete symbols: exact analysis.

$$\Psi_{\text{ref}}=\frac{|V_{\text{ref}}^2|}{|V_0^2|}, \quad (12a)$$

and the power transmission coefficient is

$$\Psi_{\text{tr}}=\frac{|V_{\text{tr}}^2|}{|V_0^2|}\frac{\text{Re}[Y_{pL}(\beta)]}{\text{Re}[Y_{pR}(\beta)]}. \quad (12b)$$

The Fourier transform of the total plate bending velocity field over the entire $x-y$ plane is given by Eq. (13a) and the radiated power can be independently calculated using Eq. (13b)

$$\overline{V}(\alpha)=\overline{V_s}(\alpha)+2\pi\delta(\alpha-\alpha_{pL}), \quad (13a)$$

$$\Pi_{\text{rad}}=2\pi k\rho c\int_{-\sqrt{1-\beta^2}}^{+\sqrt{1-\beta^2}}|\overline{V^2}(\alpha)|\text{Re}\left[\frac{N(\alpha)}{D(\alpha)}\right]d\alpha. \quad (13b)$$

Because $\text{Re}(\alpha_{pL})>1$, the incident plate velocity does not contribute to the radiated power, and Eq. (13b) may be replaced by Eq. (14a). The radiated power coefficient is given by Eq. (14b)

$$\Pi_{\text{rad}}=2\pi k\rho c\int_{-\sqrt{1-\beta^2}}^{+\sqrt{1-\beta^2}}|\overline{V_s^2}(\alpha)|\text{Re}\left[\frac{N(\alpha)}{D(\alpha)}\right]d\alpha, \quad (14a)$$

$$\Psi_{\text{rad}}=\frac{\Pi_{\text{rad}}}{V_0^2/\{2\,\text{Re}[Y_{pL}(\beta)]\}}. \quad (14b)$$

This model was used to compute power scattering coefficients for joined clamped steel plates with water on one side. The incident bending wave originated on plate L, which had half the thickness of receiving plate R. Figure 2 shows RFA results as a function of Ω_L for normal incidence ($\beta=0$), along with results previously reported in Fig. 3(a) of Ref. 6. Figure 3 shows RFA results as a function of angle of incidence φ [$\beta=\text{Re}(\alpha_{pL})\sin\varphi$] for $\Omega_L=0.3$, with comparisons to results taken from Fig. 4(a) of Ref. 3.

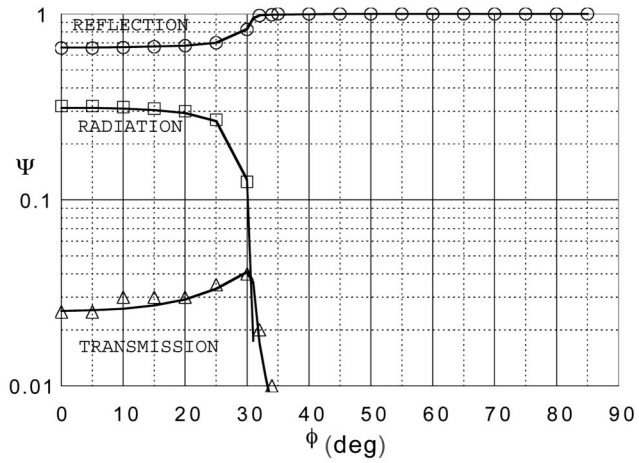


FIG. 3. Power scattering coefficients for two plates clamped along the junction (oblique incidence). Solid lines: RFA; discrete symbols: exact analysis.

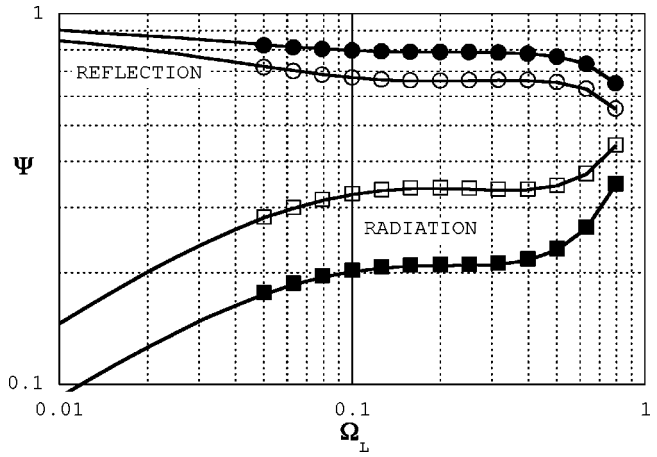


FIG. 4. Power scattering coefficients for semi-infinite plate with rigid baffle extension (normal incidence). Solid lines: RFA; discrete symbols: exact analysis. Open symbols: clamped edge; filled symbols: simply supported edge.

B. Semi-infinite plate with rigid baffle extension

If plate R is replaced by a rigid baffle, the normal velocity vanishes for $x>0$. The wave number spectrum of the scattered pressure field, J, for this condition may be obtained from Eq. (10a) by considering the mass density of plate R to become arbitrarily large. The coincidence frequency of plate R then increases without limit so that $\Omega_R \to 0$ for any finite frequency, and $\sigma_R \to 0$. As these limits are reached, $\overline{W}_R(\gamma) \to \overline{D}(\gamma)$, $W_R(\alpha) \to D(\alpha)$, and $J(\alpha)$ is given by Eq. (15)

$$J(\alpha)=\frac{V_0}{2\pi}\sum_{n=0}^{\max} B_n \frac{\alpha^n N(\alpha)}{W_L^+(\alpha)D^-(\alpha)} - \frac{V_0}{2\pi i}\frac{\rho c}{k}\frac{1}{(\alpha-\alpha_{pL}^+)}\frac{W_L^+(\alpha_{pL})D^-(\alpha_{pL})}{D(\alpha_{pL})} \times \frac{N(\alpha)}{W_L^+(\alpha)D^-(\alpha)}. \tag{15}$$

Two arbitrary constants $\{B_n\}$ are required to satisfy the boundary conditions along the edge of the plate at $x=0$. Because $[\alpha^n N(\alpha)]/[W_L^+(\alpha)D^-(\alpha)] \to \alpha^n/\alpha^4$; $|\alpha| \to \infty$, continuity of pressure allows max to equal 2, but the additional requirement of continuity of pressure derivative with respect to x further limits MAX to 1, so that the allowable coefficients are B_0 and B_1.

The UHP and LHP factors of $D(\alpha)$ are obtained from the zeros of $\overline{D}(\gamma)$. Let $\overline{D}(\pm\xi_j)=0$, $j=1, S$, where $\mathrm{Im}(\xi_j)>0$. Then

$$\overline{D}(\gamma)=b_S\prod_{j=1}^{S}(\gamma^2-\xi_j^2), \tag{16a}$$

$$D(\alpha)=b_S\prod_{j=1}^{S}[\alpha^2-(\xi_j^2-\beta^2)]. \tag{16b}$$

Because D is the limiting form of W_R, its factorization must follow the same convention, $D^-(-\alpha)=D^+(\alpha)$. Therefore, when S is an even integer, as it is for the RFA approximation used in calculations for this paper

$$D^{\pm}(\alpha)=\sqrt{b_S}\prod_{j=1}^{S}[\alpha\pm\sqrt{(\xi_j^2-\beta^2)}], \tag{17}$$

where $\mathrm{Im}[\sqrt{(\xi_j^2-\beta^2)}]>0$.

Upon combining Eq. (10b) with Eq. (15), it can be seen that the component of $V_s(\alpha)$ corresponding to the second expression on the right-hand side of Eq. (15) behaves as α^{-3} for $|\alpha|\to\infty$, and its contributions to the normal velocity and slope are each continuous across $x=0$. Therefore, if the edge of plate L is clamped (zero normal velocity and slope), both B_0 and B_1 vanish; if the plate is simply supported (zero normal velocity and bending moment), B_1 vanishes and B_0 is needed to satisfy the requirement on the second derivative of plate displacement.

Power scattering coefficients of a baffled steel plate with water on one side, for both clamped and simply supported edge conditions, are presented in Fig. 4, along with results from an exact formulation.

C. Semi-infinite plate with pressure release extension

If plate R is replaced by a pressure release surface, the pressure vanishes for $x>0$. The wave number spectrum of the scattered pressure field, J, for this condition may then be obtained from Eq. (10a) by considering the thickness of plate R to become arbitrarily small. In this limit, $\sigma_R \to \infty$, and $W_R(\alpha) \to i\varepsilon\sigma_R N(\alpha)$. The factor $i\varepsilon\sigma_R$ may be disregarded because its square root can be absorbed into the arbitrary coefficients $\{B_n\}$ in the first expression on the right-hand side of Eq. (10a), and is common to both the numerator and denominator of the second expression. The limiting form of Eq. (10a) for a pressure release surface is given by Eq. (18)

$$J(\alpha)=\frac{V_0}{2\pi}\sum_{n=0}^{\max} B_n \frac{\alpha^n N(\alpha)}{W_L^+(\alpha)N^-(\alpha)} - \frac{V_0}{2\pi i}\frac{\rho c}{k}\frac{1}{(\alpha-\alpha_{pL}^+)}\frac{W_L^+(\alpha_{pL})N^-(\alpha_{pL})}{D(\alpha_{pL})} \times \frac{N(\alpha)}{W_L^+(\alpha)N^-(\alpha)}. \tag{18}$$

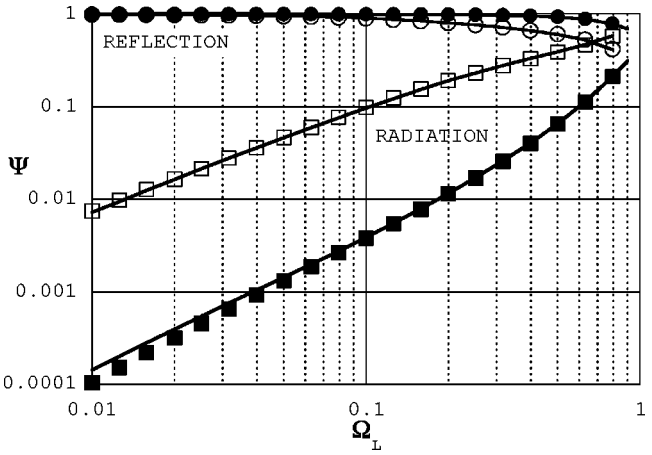


FIG. 5. Power scattering coefficients for semi-infinite plate with pressure release extension (normal incidence). Solid lines: RFA; discrete symbols: exact analysis. Open symbols: clamped edge; filled symbols: free edge.

Two arbitrary constants are again required to satisfy the boundary conditions along the edge. Because $[\alpha^n N(\alpha)]/[W_L^+(\alpha)N^-(\alpha)] \to \alpha^n/\alpha^3$; $|\alpha| \to \infty$, max may equal 1 and the allowable coefficients are B_0 and B_1. Note that the pressure derivative will be discontinuous at $x=0$. This is expected for this configuration, as discussed in the Appendix.

The UHP and LHP factors of $N(\alpha)$ are given by Eq. (19), in which $\pm\mu_j$, $j=1$, $S-1$ are the values of γ for which $\overline{N}(\gamma)=0$, $\mathrm{Im}[\sqrt{(\mu_j^2-\beta^2)}]>0$, and the factors $\pm i$ are necessary to set $N^-(-\alpha)=N^+(\alpha)$ because $S-1$ is an odd integer

$$N^{\pm}(\alpha) = \pm i\sqrt{a_{(S-1)}} \prod_{j=1}^{(S-1)} [\alpha \pm \sqrt{(\mu_j^2-\beta^2)}]. \tag{19}$$

A pressure release extension is automatically created when the same acoustic fluid is present on each side of a semi-infinite plate without a baffle extension. Power scattering coefficients of a steel plate with water on both sides, for both clamped and free-edge conditions, are presented in Fig. 5, along with results from an exact formulation.

IV. DISCUSSION

The objective of this paper was to illustrate the application of the DiPerna–Feit RFA for planar acoustic impedance to the solution of Wiener–Hopf problems for fluid-loaded plates. Figures 2–5 show good agreement between results obtained using this approximation and those obtained with an exact formulation for several representative problems.

The key feature of this formulation is that the kernel functions that appear in the integrals are polynomials that can easily be written as the product of two factors, one of which is analytic and nonzero in an upper-half plane and the other in a lower-half plane. The factorization is based on the roots of the polynomials, which are the poles of the system. Because the RFA introduces a small imaginary part to the acoustic impedance for real wave numbers in the subsonic range, the poles corresponding to the subsonic bending waves will also have small imaginary parts, even when the bending moduli are real. For the positive-going subsonic wave, the real part of the wave number is positive, but the imaginary part can turn out to be positive or negative. Regardless of the sign of the imaginary part, this pole must be associated with the UHP and its negative with the LHP.

The RFA of the acoustic impedance behaves as $|\gamma^{-2}|$ for large $|\gamma|$, while the exact impedance behaves as $|\gamma^{-1}|$. The consequence of this difference is that, based on the requirement of continuity of pressure, there appear to be three arbitrary constants for satisfying the plate boundary conditions at each edge, in contrast to the two that are needed. This issue is resolved by considering the additional requirement of continuity of pressure derivative along the plate, as discussed in the Appendix.

Assuming no power dissipation within the junction of the plates, the power scattering coefficients theoretically sum to unity, and one of the coefficients should be obtainable from the other two. It is recommended, however, that for both the exact and approximate formulations, each be calculated independently. This not only provides a check on the procedure, but prevents computational inaccuracies from distorting the results.

The illustrative examples all assumed incident subsonic bending waves, and the calculations were restricted to $\Omega_L < 1$ because the RFA pole corresponding to the subsonic wave cannot be uniquely identified above the coincidence frequency. If plate R is thicker than plate L, so that $\Omega_L < 1$ but $\Omega_R > 1$, only the reflection and radiation coefficients can be directly determined.

Because the formulation is in terms of polynomials, and evaluation of integrals is in terms of residues, quantitative solutions are much faster than for the exact analysis. Therefore, if power scattering coefficients are needed within a larger problem, they can be calculated as required with known parameters, rather than precalculated for a set of expected parameters and accessed by table lookup, as had been suggested in Ref. 6.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under the Surface Ship Flow Noise Task: Mr. Stephen C. Schreppler, Program Monitor. I am grateful to my colleague Dr. Daniel DiPerna for providing the RFA coefficients listed in Table I and for helpful comments regarding the analysis and manuscript.

APPENDIX: CONTINUITY OF PRESSURE DERIVATIVE

If the normal velocity is finite and continuous for all x, its Fourier transform must behave as $|\alpha|^{-1-\nu}$; $\nu>0$ for large $|\alpha|$. Because the exact acoustic impedance behaves as $|\alpha|^{-1}$ for large $|\alpha|$, the Fourier transform of the pressure behaves as $|\alpha|^{-2-\nu}$, and both the pressure and its derivative along the surface are continuous functions of x. Therefore, the pressure derivative is continuous for the two joined semi-infinite plates, and for the semi-infinite plate with rigid baffle extension when it is clamped or simply supported at the baffle.

The situation is different when the velocity is discontinuous, as it is for a rigidly extended plate with a free edge.

As an illustration, consider the simpler case of the normal velocity distribution $v(x)=V$, $x<0$; $v(x)=0$, $x>0$. The resulting pressure is given by Eqs. (A1)

$$P(x)=\frac{\rho cV}{2}\left[1+k\int_0^{-x}H_0^{(1)}(ky)dy\right];\quad x<0, \tag{A1a}$$

$$P(x)=\frac{\rho cV}{2}\left[1-k\int_0^{+x}H_0^{(1)}(ky)dy\right];\quad x>0. \tag{A1b}$$

The pressure is continuous for all x. The pressure derivative is given by Eqs. (A2)

$$\frac{dP(x)}{dx}=-\frac{\rho\omega V}{2}H_0^{(1)}(-kx);\quad x<0, \tag{A2a}$$

$$\frac{dP(x)}{dx}=-\frac{\rho\omega V}{2}H_0^{(1)}(+kx);\quad x>0. \tag{A2b}$$

The pressure derivative is an even function of x and can be considered continuous, except for a very weak logarithmic singularity at $x=0$. The RFA approximation maintains continuity and ignores the singularity.

Finally, consider a semi-infinite plate with a pressure release extension. A simple example is given by the conditions $v(x)=V$, $x<0$; $p(x)=0$, $x>0$. A Wiener–Hopf analysis using the exact acoustic impedance can be used to obtain the pressure

$$p(x)=\rho cV-\frac{\rho cV}{2\pi i}\sqrt{k}\int_{-\infty}^{+\infty}\frac{\exp(i\alpha x)}{(\alpha-i0^+)}\frac{d\alpha}{\sqrt{(\alpha+k)}}, \tag{A3a}$$

$$p(x)=\rho cV\left[1-\frac{\sqrt{k}}{\pi}\int_k^{\infty}\frac{\exp(-i\alpha x)}{\alpha\sqrt{(\alpha-k)}}d\alpha\right];\quad x<0, \tag{A3b}$$

$$p(x)=0;\quad x>0. \tag{A3c}$$

The pressure is continuous at $x=0$. The derivative of the pressure is given by Eqs. (A4)

$$\frac{dp(x)}{dx}=\frac{i\rho cV}{\pi}\sqrt{k}\int_k^{+\infty}\frac{\exp(-i\alpha x)}{\sqrt{(\alpha-k)}}d\alpha$$

$$=\frac{\rho\omega V}{\sqrt{(-\pi kx)}}\exp[-i(kx-\pi/4)];$$

$$x<0, \tag{A4a}$$

$$\frac{dp(x)}{dx}=0;\quad x>0. \tag{A4b}$$

In this case, the pressure derivative is discontinuous at $x=0$, with a weak algebraic singularity as $x=0$ is approached through negative values. The RFA approximation maintains the discontinuity but ignores the singularity.

[1] M.S. Howe, *Acoustics of Fluid–Structure Interactions* (Cambridge University Press, New York, 1998), Sec. 4.4.

[2] A.N. Norris and G.R. Wickham, "Acoustic diffraction from the junction of two flat plates," Proc. R. Soc. London, Ser. A **451**, 631–655 (1995).

[3] D.A. Rebinsky and A.N. Norris, "Acoustic and flexural wave scattering from a three-member junction," J. Acoust. Soc. Am. **98**, 3309–3319 (1995).

[4] I.D. Abrahams, "The application of Padë approximants to Wiener–Hopf factorization," IMA J. Appl. Math. **65**, 257–281 (2000).

[5] D.T. DiPerna and D. Feit, "An approximate Green's function for a locally excitated fluid-loaded thin elastic plate," J. Acoust. Soc. Am. **114**, 194–199 (2003).

[6] M.L. Rumerman, "Estimation of broadband power levels radiated from turbulent boundary layer-driven ribbed plates having dissimilar sections," J. Acoust. Soc. Am. **114**, 737–744 (2003).

The effects of ignored seabed variability in geoacoustic inversion

Anna-Liesa S. Lapinski[a] and David M. F. Chapman
Defence Research and Development Canada—Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 3 November 2004; revised 24 March 2005; accepted 1 April 2005)

Geoacoustic inversion using a matched-field inversion algorithm is a well-established technique for estimating the geoacoustic parameters of the seabed. This paper demonstrates how parameter estimation can be affected by unknown or wishfully ignored random range dependence of the true environment when the inversion model is—for practicality—assumed to be range independent. Simulations with controlled statistics were carried out using a simple shallow water model: an isospeed water column over a homogeneous elastic halfspace. The inversion parameters included water depth, compressional speed in the seabed, seabed density, and compressional wave attenuation. On average the environment is range independent: some parameters are constant while other parameters are random with range-independent means and variances. A Parabolic Equation underwater acoustic propagation model is used to calculate the simulated data fields for the range-dependent environment as well as to calculate the model fields for the range-independent inversion model. The Adaptive Simplex Simulated Annealing inversion algorithm is used to estimate the best-fit solution. It is found that ignoring the variability of even a single geoacoustic parameter leads to significant and correlated uncertainty (bias and variance) in the estimation of all inverted parameters. Results are presented for range variation of compressional sound speed and water depth. [DOI: 10.1121/1.1921449]

PACS numbers: 43.30.Pc, 43.60.Kx [AIT] Pages: 3524–3538

I. INTRODUCTION

Geoacoustic inversion using a matched-field inversion (MFI) algorithm is a well-established technique for estimating the geoacoustic parameters of the seabed.[1–9] This paper demonstrates how parameter estimation can be affected by unknown or wishfully ignored random range dependence of the true environment when the inversion model is assumed to be range independent. A simple shallow water model is used for simulations with controlled statistics: the inversion parameters include water depth, seabed compressional speed, seabed density, and compressional wave attenuation. All environmental parameters are either assumed to be constant in range or random with range-independent means and variances. A Parabolic Equation underwater acoustic propagation model[10] is used to calculate both the simulated data fields for the range-dependent environment as well as the model fields for the range-independent inversion. It is found that ignoring the variability of even a single geoacoustic parameter leads to significant and correlated uncertainty (bias and variance) in the estimation of all inverted parameters.

Geoacoustic inversion is a technique used to determine geoacoustic properties in the ocean. When acoustic fields interact with the ocean environment, including the ocean bottom, they are influenced by properties of the environment. In the technique of geoacoustic inversion, measured geoacoustic fields **d** are used to determine the M unknown parameters of a postulated model $\mathbf{m}=\{m_i, i=1,M\}$ that represents unknown geoacoustic properties of the ocean that have influenced the measured field. The determination of **m** using **d** is the solution to the inverse problem.

The proper formulation of the inverse problem is imperative. If it is not formulated appropriately then the results of the inversion could be meaningless. In particular, determining a suitable model parametrization is an important preceding step to applying an inversion algorithm. *A priori* information about the ocean environment can help determine an appropriate model parametrization. When adequate information about the ocean environment is not available, techniques such as the underparametrized approach (e.g., as applied in Lapinski and Dosso[11] as well as Refs. 1 and 2) and the overparametrized approach used in conjunction with the inversion algorithm can be employed to determine the best model parametrization as well as **m**. In this research, however, the goal was not to determine the best parametrization but rather to observe the effects of assuming a particular parametrization.

The research presented here is a documentation of some of the results from a single formulation of an inverse problem. It presents the results obtained by assuming that the unknown ocean bottom properties could be modeled as if they did not vary with range (i.e., they are range independent) while in reality the properties were randomly range dependent. Assuming the properties of the ocean environment can be treated as range independent is an appealing assumption in geoacoustic inversion because it can simplify the model parametrization and thereby can simplify the inversion. This paper considers how ocean bottom parameter estimation is affected by unknown or wishfully ignored random variation in certain environments.

[a]Electronic mail: liesa.lapinski@drdc-rddc.gc.ca

Geoacoustic inversion is a nonlinear inverse problem. It has no direct solution. The technique of matched-field inversion is commonly used to ascertain a solution. In MFI, measured acoustic fields **d** are compared with replica (modeled) acoustic fields **d**(**m**) that have been generated using an acoustic propagation model with a proposed model **m**. Typically, the goal of MFI is to minimize the objective function $E(\mathbf{m})$. To evaluate the proposed parameter values of the model **m** an error function is defined that quantifies the mismatch between the measured field **d** and the modeled field **d**(**m**). In this investigation, the objective function $E(\mathbf{m})$ is the same as the error function. The error function used in this research is described in Sec. II B. Minimizing the objective function requires searching for the model that has the smallest mismatch E. Before the search begins, upper and lower bounds must be placed on each model parameter $m_i^- < m_i < m_i^+$. This defines the parameter space that is used in solving the inverse problem. The parameter space is then searched using an MFI algorithm to find the best model **m**. The search can be a sizable activity. In geoacoustic inversion, many different MFI algorithms have been applied with varying success and efficiency.[1–8]

Given the variability that exists in the geoacoustic properties of the oceans, studying how well geoacoustic inversion techniques can be applied to range-dependent environments is a worthwhile endeavor. For example, in 2001 the Geoacoustic Inversion Techniques (GAIT) workshop was held to provide participants with an opportunity to apply their geoacoustic inversion techniques to range-dependent test cases.[9] The papers detailing the results of this workshop (including Ref. 9) are contained in a special issue of the *IEEE Journal of Oceanic Engineering* on geoacoustic inversion in range-dependent shallow-water environments. One reason to investigate the effect of ignoring range dependence is because if the range dependence can be ignored, less computationally intensive geoacoustic models can be employed in the inversions. Similarly, for the GAIT workshop data, Nielsen *et al.*[12] found that it was appropriate to apply both a Parabolic Equation (PE) based acoustic model (RAMGEO) and a (less-accurate; i.e., less computationally intensive) adiabatic model (PROSIM) to the data from the test case referred to as TC1. This could be done because the environment is only weakly range dependent: the environment is a smooth shallow-sloped downslope environment. The test case environment in TC3 is a simulation of an uplifted fault that consists of a rectangular intrusion of the basement material in the sediment layer. For TC3 the adiabatic model was thought to have failed due to the intrusion in the environment that caused significant mode coupling.[12] Only the PE acoustic model RAMGEO could be applied in this case. Also in that special issue, Harrison and Siderius[13] applied the ''effective'' water depth technique to two of the workshop test cases, the benefit being that a range-dependent environment can be replaced by a simplified equivalent range-independent environment for faster inversions. In other literature, Siderius *et al.*[14] safely neglected range dependence using a towed horizontal line array (HLA) and a towed source which in turn allowed for the neglect of range dependence during the inversions while still allowing them to map a range-dependent environment.

In this paper the Adaptive Simplex Simulated Annealing (ASSA)[1] MFI algorithm is applied to synthetic geoacoustic data generated for an ocean environment that has one ''unknown'' ocean bottom property that varies with range. During the inversions a range-independent model is assumed. While studying the effect of ignoring range dependence, effects of simply applying the inversion algorithm should be kept in mind. Fallat *et al.*[15] investigated algorithm-induced variability for three geoacoustic inversion techniques, including ASSA. It was found that ASSA produced the lowest (i.e., best) objective function values and the lowest variability between inversion solutions for all environments it was applied to. However, it was noted that algorithm-induced variability should be evaluated on a case-by-case basis since the variability varied between test environments.

The goal of this paper is to investigate some of the effects of assuming a range-independent environment when performing an inversion. Section II introduces the environment, discusses ASSA and the error function used in this work, and introduces PE acoustic propagation models and the propagation model used in this investigation. Section III presents the results of the studies done for this investigation, which include results for an environment with range-dependent sediment compressional speed and for an environment with range-dependent water depth. Section IV summarizes the conclusions made.

II. SETUP: MODELS AND ERROR FUNCTION

In the following the general environment is described, the ASSA algorithm is introduced, the error function used with the algorithm is discussed, and the acoustic propagation model used in this investigation is introduced.

A. Environment

In order to look at the effect of ignoring range dependence when doing inversions, a prescribed ocean environment and ''experimental setup'' were first devised. The synthetic ocean environment used in this investigation is simple but realistic inasmuch as it is based on a model of the Sable Island Bank, which is part of the Scotian Shelf off the east coast of Canada where the ocean bottom consists of three general layers: sand and gravel over till over bedrock.[16] The simplified version of this environment used for this investigation, as seen in Fig. 1, consists of a water column over a semi-infinite layer of sand where water sound speed is $V_w = 1460\,\text{m/s}$,[17] water depth is $H = 80\,\text{m}$,[18] sediment density is $\rho = 2.06\ \text{g/cm}^3$,[17] sediment attenuation coefficient is 0.26 dB/km Hz,[17] which gives an attenuation of $\alpha_p = 0.436\,\text{dB}/\lambda$, shear speed is $V_s = 260\,\text{m/s}$,[19] and shear attenuation is calculated to be $\alpha_s = 2.6\,\text{dB}/\lambda$.[20] The sediment compressional speed is taken as $V_p = 1677\,\text{m/s}$, the average of calculated sound speeds in Ref. 18 for the southern part of their survey area. That average has a standard deviation of ± 71 m/s, which put a maximum limit on how much the compressional speed could be allowed to vary in this investigation. The sediment compressional speed is realistic for

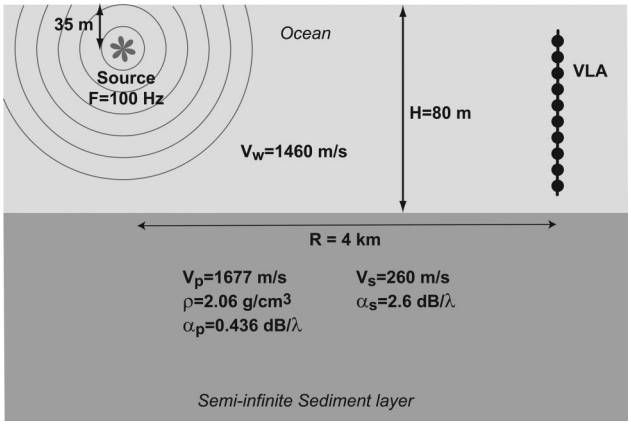


FIG. 1. Schematic representation of the basic ocean environment and "experimental setup" used for this study, including source frequency F, water sound speed V_w, water depth H, range R, sediment compressional speed V_p, sediment density ρ, compressional wave attenuation α_p, shear speed V_s, and shear attenuation α_s.

the Sable Island Bank in that it falls between general compressional speed values for "Sable Island Sand and Gravel."[19]

A single source at 35 m depth and a vertical line array (VLA) with receivers every meter from 5 m depth to 75 m depth were used when generating the acoustic fields (Fig. 1). Typically, the source frequency was $F=100$ Hz and the range, R, from the source to the VLA was 4 km. In addition, to reduce model artifacts, such as unrealistic reflections from where the ocean bottom is terminated in the acoustic model, an absorbing bottom was included.

B. Matched-field inversion

Optimizing through MFI methods requires the quantification of the difference between the measured and modeled data through an error function in order to evaluate the fitness of the model. The following starts by describing the error function followed by a brief description of the optimization algorithm applied here.

The essence of matched field techniques is that a field generated using a proposed model is compared to the actual measured field and then assessed. The difference between the fields (the fitness or mismatch) requires quantification using an error function. The general form of the error function $E(\mathbf{m})$ used in this work for nF source frequencies and nS data snapshots (i.e., the number of data sets being used simultaneously in the inversion at source frequency F) for observed data $\mathbf{d}=\{\mathbf{d}_{ij};i=1...nF;j=1...nS\}$ is

$$E(\mathbf{m})=\sum_{i=1}^{nF}\sum_{j=1}^{nS} B_{ij}(\mathbf{m})|\mathbf{d}_{ij}|^2/v_{ij}, \tag{1}$$

where v_{ij} is the data variance and $B_{ij}(\mathbf{m})$ is the normalized incoherent Bartlett mismatch. The normalized incoherent Bartlett mismatch has the formulation

$$B_{ij}(\mathbf{m})=1-\frac{|\mathbf{d}_{ij}^{\dagger}(\mathbf{m})\mathbf{d}_{ij}|^2}{|\mathbf{d}_{ij}|^2|\mathbf{d}_{ij}(\mathbf{m})|^2}, \tag{2}$$

which can vary between 0 and 1 where 0 represents a perfect match (i.e., parallel vectors) and 1 represents a perfect mismatch (i.e., orthogonal vectors). Equation (1) is a form of the incoherent Bartlett processor. (It is more common to see it without the summation over data snapshots, for example as derived in Ref. 21.) The incoherent Bartlett processor is a common error function (also mismatch calculation) used for MFI of spatial acoustic-field data. It is a likelihood based processor derived when source amplitude and phase at each frequency are unknown. For a list of several other likelihood based processors see Refs. 21 and 22. The maximum likelihood model is found by minimizing Eq. (1) given the simplifying assumption that the measured data have complex, zero-mean, Gaussian-distributed random variable errors that are uncorrelated across space, frequency, and time, with identical variance at each array sensor.

Only relative values of the error function are important for optimization. If it can be assumed that $|\mathbf{d}_{ij}|^2/v_{ij}$ is constant with respect to frequency and snapshot then the error function [Eq. (1)] can be rewritten as

$$E(\mathbf{m})=\frac{1}{nF\cdot nS}\sum_{i=1}^{nF}\sum_{j=1}^{nS} B_{ij}(\mathbf{m}), \tag{3}$$

which varies on [0,1]. Equation (3) is the error function used in this work.

The results contained in this paper, as with any data modeling investigation, depend in part on the particular error function used. It is possible that alternative results could have been obtained using a different error function. For example, Dosso and Wilmut[21] did a study on the effects of using different levels of information about the source amplitude and phase in MFI. They compared results obtained using three maximum likelihood processors and found that smaller parameter uncertainties were achieved if source information was used. The error function Eq. (3), which uses no source information, was chosen because often in MFI source spectrum information is not available and therefore these results would be of greater use. While the details of the results might change if a different error function was used, it would be expected that the general conclusions made here remain the same.

To perform the investigation reported in this work, the hybrid inversion algorithm ASSA[1] was used to minimize Eq. (3). A hybrid inversion method combines a local with a global inversion method in order to combine the strengths and alleviate the weaknesses of the two original methods. In this case, the local method of Downhill Simplex (DHS)[23,24] is combined with the global method of Simulated Annealing (SA).[25,24] The DHS component helps the algorithm converge upon an estimate of the global minimum in an efficient manner while the SA prevents the algorithm from getting stuck in suboptimal minima, and allows the parameter space to be widely searched. It has been shown to be an effective optimization algorithm[1,11] and has been applied to data from environments that have both range-independent[1] and range-dependent[11] properties. The reader is directed toward Ref. 1 if a description of the algorithm is required.

ASSA is an optimization algorithm. The error function Eq. (3) is used to help the algorithm evaluate the proposed

parameter models, **m**, and subsequently ASSA minimizes Eq. (3). The model evaluation process in ASSA comes from the SA component of the algorithm. SA is based on the process of crystal growth through thermodynamic annealing. In the SA algorithm, the system (model) is repeatedly perturbed and the perturbations are accepted or rejected based on whether the perturbed system satisfies the METROPOLIS criterion[26]

$$\xi \leqslant \exp[-\Delta E/T], \tag{4}$$

where ΔE is the difference between the energies (mismatches) of the proposed and current states (models), T is temperature (a control parameter) that decreases throughout the algorithm, and ξ is a uniform random number on [0,1]. Any perturbed system (model) that achieves a lower energy (mismatch) or in other words moves downhill in the objective function is accepted while a perturbed system (model) with a higher energy (i.e., moves uphill) may still be accepted. ASSA uses the METROPOLIS criterion to accept or reject perturbed models and for this investigation the mismatch, E, was calculated using the error function Eq. (3). To ensure that the best model possible is found, the final phase of the algorithm is a quenching that applies the pure DHS algorithm using the current best model found by applying ASSA. Without going into more algorithm detail, this is the underlying technique used to find the model with the lowest mismatch.

A detailed description of ASSA has not been presented here, but the values of certain ASSA parameters should be stated for the sake of the experimental setup used in this work. The ASSA control parameters mentioned here are discussed more fully in Refs. 1 and 2. One set of control parameters, the annealing schedule, consists of the starting temperature T_0 [where the temperature T is the parameter defined in Eq. (4)], the rate of reducing the temperature β ($\leqslant 1$), the maximum number of temperature steps, and the number of accepted perturbations at each temperature step. (The actual number of temperature steps varied with each application of ASSA because a convergence criterion was enforced prior to quenching so that computation time was not wasted.) In general, the larger the annealing parameter values are the slower and more extensive the exploration of the search space is. The difficulty of the inversion problem defines how intensive a search should be conducted. Some initial runs with different annealing schedules are usually required to find a schedule that generates results that cannot be appreciably improved by intensifying the search. The annealing schedule used in this work was typically $T_0 = 0.7$, $\beta = 0.975$, 5 accepted perturbations, and a maximum of 400 temperature steps. In certain studies, when it was noted that the algorithm was having problems converging prior to quenching, the annealing schedule had to be altered: the maximum number of temperature steps went up to 600 and the number of accepted perturbations reached 7. The changes intensified the search. In addition to these parameters, the Cauchy distribution parameters[1] S and s, which influence the adaptive nature of the algorithm, were set to $S = 30$ and $s = 3$.

C. Setup of the propagation model PECan

Matched-field inversion requires the calculation of the acoustic fields at the receivers given the proposed model in order to evaluate the model using an error function. There are a variety of types of acoustic propagation models that can be used to calculate acoustic fields. The PE method is the method used here because it can be used to calculate fields in both range-independent and range-dependent environments. When calculating fields for range-dependent environments the PE method can be less computationally intensive than some other methods, such as the coupled-modes method. The particular propagation model used here is PECan.[10] This section describes the problem-specific setup of PECan along with a feature of PECan that was utilized to help generate the continuously variable properties.

PECan was used to calculate the ''measured'' fields **d** and the replica fields **d**(**m**) needed to evaluate the model. While a range-independent environment was assumed during the inversions, to calculate **d** an acoustic model that could handle range dependence was also needed. There can be differences between an acoustic field calculated using different acoustic propagation models. The significance of these differences depends on what the calculated acoustic field is used for. To minimize consequential numerical errors between the ''measured'' and replica fields, i.e., numerical differences that could affect conclusions, PECan was used to calculate both the fields. Due to the far-field approximation made in PE models, the VLAs were also placed well in the far-field, several kilometers from the source.

The accuracy and computation time of a PE model can be affected by choices made during the setup of the model. To solve a parabolic equation, the environment is divided laterally (Δr) and vertically (Δz). Using a split-step Padé solution method, larger range steps can be used compared to other PE solution methods thereby reducing the computation time.[10] Finding the appropriate grid size to use when calculating an acoustic field is an important step. Finer grids provide a more accurate solution but require larger computations. For inversion purposes it is therefore important to strike a balance between grid size and the accuracy needed. A similar time versus accuracy balance is needed for the feasible number of Padé expansion terms (N_P): computations increase as the number of expansion terms (accuracy) increase.

PECan has many options and parameters that must be set prior to applying the model. For the inversions in this investigation an analytic Green's starting field, the energy-conservation option, and the split-step Padé solution method were used. The PECan parameters that needed to be defined for the inversions were the values of the grid spacing Δz and Δr, the depth at which the absorbing bottom starts H_a, and the number of Padé terms N_P. The thickness of the absorbing bottom ΔH_a also needed to be determined.

To determine the appropriate PECan parameter values one realization of an ocean environment based on Fig. 1 with variable compressional speed $v_p(x)$ was created (more details on how variability was introduced are discussed in the following section). The choice of value for Δz, Δr, H_a, ΔH_a and N_P contributes to how computationally intensive

TABLE I. Results of PECan parameter studies. For 25, 50, 100, and 200 Hz data the chosen Δz, Δr, N_P, H_a (H_a is a depth measured from the ocean surface) and ΔH_a are included.

Frequency (Hz)	Δz (m)	Δr (m)	N_P	H_a (m)	ΔH_a (m)
25	1.0	40	3	400	100
50	0.4	50	3	140	50
100	0.25	8	3	110	100
200	0.125	10	4	90	10[a]

[a]Some inversions were done with a $\Delta H_a = 100$ m.

each call to PECan will be during the inversions. Appropriate values for each parameter were determined in the following way: A ''true'' field was first generated using a computationally intensive set of PECan parameter values. Without changing the environment, with the exception of the artificial absorbing layer, PECan was repeatedly run using all possible combinations of a selection of PECan parameter values. The calculated (replica) fields were compared to the originally generated ''true'' field using the error function in Eq. (3) with $nF=1$ and $nS=1$. From these results, a set of PECan parameter values that gave an acceptable mismatch and an acceptable computational time was chosen. An acceptable mismatch was a comparative value based on the best mismatches achieved for the PECan parameters combinations. A mismatch less than 10^{-4} was acceptable but a mismatch close to machine precision was desirable. An acceptable computational time for the forward propagation model was again a comparative value based on the computation times for each combination of PECan parameter values. On a PC with a 2 GHz CPU, a few seconds or less was desirable. The process of determining the best parameter values had to be repeated for each source frequency used in this investigation, which included 25, 50, 100, and 200 Hz. The PECan parameter values picked for each frequency are quoted in Table I.

Another of PECan's features was utilized to help create the continuously variable ocean environments used in this investigation when calculating **d**. In PECan the environment is defined using environmental profiles. PECan allows for linear interpolation between environmental profiles. The range-dependent properties for environmental profiles with range variability were specified at evenly spaced range points, and PECan linearly interpolated between each profile as illustrated for variable compressional speed in Fig. 2. The linear interpolation feature was employed so that the variability was continuous with range within the limit of the grid spacing.

III. RESULTS

In this section the method of investigation is explained and the results are presented. The inversion results have been grouped into studies. For each study the ''experimental setup'' was established and the range-dependent variability had a prescribed mean and standard deviation. Typically, 100 inversions were performed per study. Unless stated otherwise, the true environment used to generate the ''measured'' field **d** had a different realization of the range-dependent variability for each inversion in the study. This corresponds to doing inversions using data from multiple tracks for a certain ocean environment. The spread and correlations of the optimal model parameter solutions found through the inversions could then be examined and compared within and between the studies.

When ASSA was applied it was assumed that the ''unknown'' ocean bottom properties could be modeled as range independent. A four-parameter model was inverted for in the inversions. The parameters represented water depth H, sediment compressional speed V_p, density ρ, and attenuation α_p. The shear speed V_s and attenuation α_s were assumed to be known and constant. When generating the ''measured'' field **d**, prior to the inversion, the true values of the model parameters were those quoted in Fig. 1, except, of course, for the property that varied with range in the particular study. The results presented in the following sections include results for variable compressional speed and variable water depth.

The shear parameters were held constant to reduce the dimensionality of the problem. Even without environmental variability, there is ambiguity among the geo-acoustic parameters of this simple model.[27] It will be shown that when shear parameters are held constant, the inversion (when successful) consistently finds correct parameter values when the environment does not vary. This provides a good baseline case for reference purposes, with no inherent ambiguity. More complex cases of multiple range-dependent parameters could be a topic of further study, using the current work as a starting point.

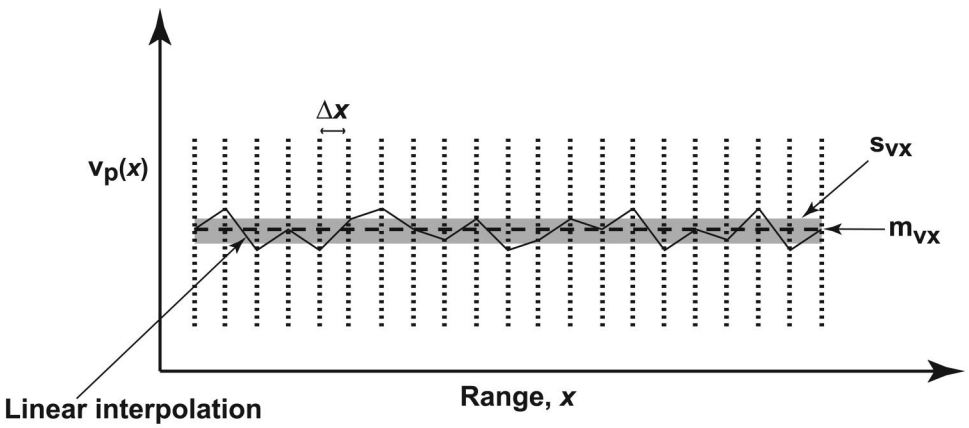


FIG. 2. Schematic representation of the compressional speed variability. The horizontal axis represents range x and the vertical axis represents the variable compressional speed $v_p(x)$. The dotted vertical lines represent where the PECan environmental profiles are defined, which is every Δx interval. PECan linearly interpolates between where the fluctuations are defined. The resulting linearly interpolated range-dependent profile has a set mean m_{vx} and standard deviation s_{vx}.

A. Variable compressional speed

The sediment compressional speed was one of the ocean bottom properties that was allowed to vary with range $v_p(x)$ in this investigation. To create $v_p(x)$, fluctuations were generated using scaled pseudorandom numbers picked from a uniform distribution. For $N+1$ environmental profiles defined in PECan (i.e., N segments), $N+1$ fluctuations were generated to form a set of fluctuating compressional speed values $\{[v_p]_i, i=0,N\}$. These values were assigned to the $N+1$ environmental profiles and PECan was allowed to linearly interpolate between the profiles to create $v_p(x)$ (Fig. 2). For variable compressional speed, the environmental profiles were 100 m apart. Fluctuations were generated so that $v_p(x)$ would have a set (true) mean $m_{vx}=1677$ m/s and standard deviation s_{vx} as defined by

$$m_{vx}=\frac{1}{2N}\sum_{i=0}^{N}([v_p]_{i+1}+[v_p]_i), \tag{5}$$

$$s_{vx}=\pm\sqrt{-m_{vx}^2+\frac{1}{3N}\sum_{i=0}^{N}([v_p]_{i+1}^2+[v_p]_i^2+[v_p]_{i+1}[v_p]_i)}. \tag{6}$$

For a derivation of Eqs. (5) and (6) please refer to Appendix A.

Table II summarizes the features of each study presented here for variable compressional speed (studies V1–V8). Four studies include inversions for 100 Hz source frequency (V1–V4) while the others include source frequencies of 25, 50, and 200 Hz (V5, V6, and V7, respectively) and multiple source frequencies 100 and 200 Hz (V8). For some numerical propagation models, such as the one used here, the computational effort to calculate the field prohibitively increases with increasing frequency. Studies using frequencies higher than 200 Hz were not attempted due to the computation time that inversions at those frequencies require. For all studies except study V8, $nF=1$ and $nS=1$ in Eq. (3). For study V8, $nF=2$.

TABLE II. Summary of characteristics of each study presented in this paper. The data include the study number, the number of realizations of the ocean bottom used N_R, the number of inversions performed in the study N_I, the number of source frequencies used N_F, the source frequencies used F, the range between source and VLA R, and standard deviations of compressional speed variability s_{vx} and water depth variability s_{hx}.

Study number	N_R	N_I	nF	F (Hz)	R (km)	s_{vx} (m/s) for $v_p(x)$
A	1	29	1	100	4	0
V1	100	100	1	100	4	±5
V2	100	100	1	100	4	±20
V2(a)	1	31	1	100	4	±20
V2(b)	1	31	1	100	4	±20
V3	100	100	1	100	4	±40
V4	100	100	1	100	15	±20
V5	100	100	1	25	4	±20
V6	100	100	1	50	4	±20
V7	100	100	1	200	4	±20
V8	100	100	2	100, 200	4	±20
Study number	N_R	N_I	nF	F (Hz)	R (km)	s_{hx} (m) for $h(x)$
H1	100	100	1	100	4	±0.125
H2	100	100	1	100	4	±0.250
H3	100	100	1	100	4	±0.500
H4	100	100	1	100	4	±0.750
H5	100	100	1	100	4	±1.000

For study V1 (Table II), the true environments had a range-dependent compressional speed $v_p(x)$ with a standard deviation of $s_{vx}=\pm 5$ m/s about a mean $m_{vx}=1677$ m/s. Figures 3(i)–3(l) are plots of the mismatches of the 100 inversion solutions versus the model parameter results. As can be seen in Figs. 3(i)–3(l), solutions that achieved a mismatch below $\sim 6\times 10^{-4}$ are clustered about the true mean m_{vx} or true values used to generate the environments. However, for all four model parameters, if the inversion result had a mismatch greater than $\sim 6\times 10^{-4}$ a wider distribution of results not centered about the true values or true mean value was obtained.

For comparison, Fig. 4 is a plot comparable with Figs. 3(i)–3(l). It too is a plot of mismatch versus inversion result; however, for these inversions the true environment was actually a range-independent environment as shown in Fig. 1 (study A, Table II). Twenty-nine inversions were performed, and most of the inversion results (21 of them) were tightly clustered about the true parameter values. These 21 results all had mismatches of $\sim 10^{-7}$; in other words the mismatch achieved 0 for the given machine precision. The remaining inversion results converged to suboptimal solutions and were not scattered about the true values. This figure shows how tight a cluster about the true solution can be achieved and that the inversion algorithm does not always find the best solution. A possible reason for the latter observation is that the inversion algorithm, as it was set up for this investigation, did not search the search space as thoroughly as it should have. Therefore, for the data shown in Figs. 3(i)–3(l) there are two possible reasons for the suboptimal results: for those environments either no significantly better model result existed (i.e., a result with a significantly lower mismatch did not exist) or the algorithm's search was unsuccessful. Regarding the latter possibility, inversion results in this study are not guaranteed to be optimal, despite having acceptable control parameters. Nevertheless, the general trends which lead to the conclusions of this study are still evident even in the presence of a small segment of outliers, which are less

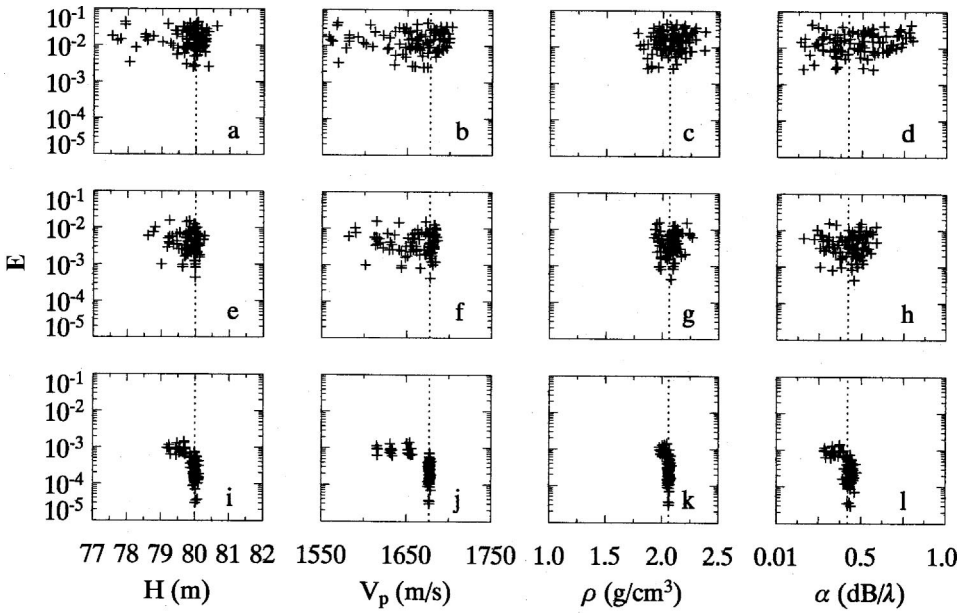


FIG. 3. Mismatch vs model parameter values for the solution sets of studies V3 ($s_{vx}=\pm 40$ m/s) are shown in (a)–(d), V2 ($s_{vx}=\pm 20$ m/s) are shown in (e)–(h), and V1 ($s_{vx}=\pm 5$ m/s) are shown in (i)–(l). The model parameters include water depth H, compressional speed V_p, density ρ, and attenuation α. For H, ρ, and α the vertical dotted lines represent the true model parameter values used to generate the measured acoustic field, and for V_p it represents the m_{vx}. The abscissa limits represent the bounds used in the inversions.

noticeable in the range-dependent environments.

Figures 3(a)–3(d) and 3(e)–3(h) show results for studies with variable compressional speed that have standard deviations of $s_{vx}=\pm 20$ m/s [study V2, Figs. 3(e)–3(h)] and $s_{vx}=\pm 40$ m/s [study V3, Figs. 3(a)–3(d)]. As can be seen, as the variability increases, the average mismatch increases as does the spread of solutions. It is likely that the average mismatches increase because there is no model that exists in the search space that can generate a replica field that is close to the "measured" field. Looking at Figs. 3 and 4, it appears that as more variability is added, the depth of the global minimum for all environments shrinks until there is no more global minimum at the true parameter values and the true mean m_{vx}.

To again look at how predictable the results of the inversions were, two sub-studies were done for study V2 [see Table II, studies V2(a) and V2(b)]. Two realizations of the ocean environment were chosen out of the original V2 study, and multiple inversions were performed on the fields generated from those environments. The results of those inversions are found in Fig. 5. As can be seen, for the environment that produced the results in Figs. 5(a)–5(d) there appeared to be a global minimum in the objective function at the true mean m_{vx} and true parameter values, however for some of the inversions a good estimate of the global minimum was not located. For the environment represented by Figs. 5(e)–5(h) there does not appear to be a well-defined global minimum. Except for density ρ, a spread of values at approximately the same mismatch was found. Therefore, for an environment that does not have a well-defined global minimum the result from a single inversion will likely be just one of many equally good solutions. The environment for study V2(b) [Figs. 5(e)–5(h)] could lack a well-defined optimal solution but the environment for study V2(a) [Figs. 5(a)–5(d)] seems to have a well-defined optimal solution.

The spread and concentrations of the $s_{vx}=\pm 5$ m/s, ± 20 m/s, and ± 40 m/s data (studies V1, V2, and V3) are better seen in Figs. 6(j)–6(l), 6(e)–6(h), and 6(a)–6(d), respectively. Figure 6 presents histogram plots of the results for each study. If one assumes a range-independent environment while performing an inversion, it can be seen that for small variations, like $s_{vx}=\pm 5$ m/s, one is likely to obtain a solution close to the true and true mean values. In Figs. 6(j) and 6(k) a high concentration of the model parameter values were clustered tightly about m_{vx} and the true ρ value, respectively. It can be concluded that for the general environment used in this investigation, if the variability is small, one is likely to find a solution that represents the true environment after performing only one inversion. As the variability increases, the solutions found by single inversions of the environment are increasingly unreliable estimates of the true environment. This is reminiscent of adding increasing amounts

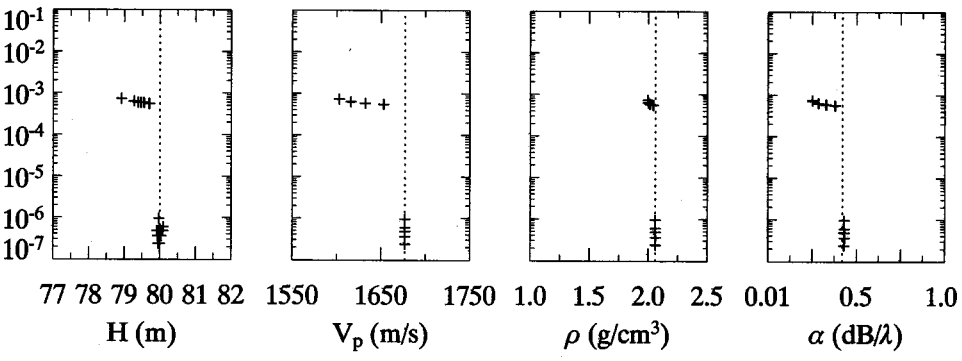


FIG. 4. Mismatch vs model parameter values for the solution set of study A. The true environment for this study is range independent (i.e., no parameter fluctuations have been added). The model parameters, abscissa limits, and vertical dotted lines are as described in Fig. 3.

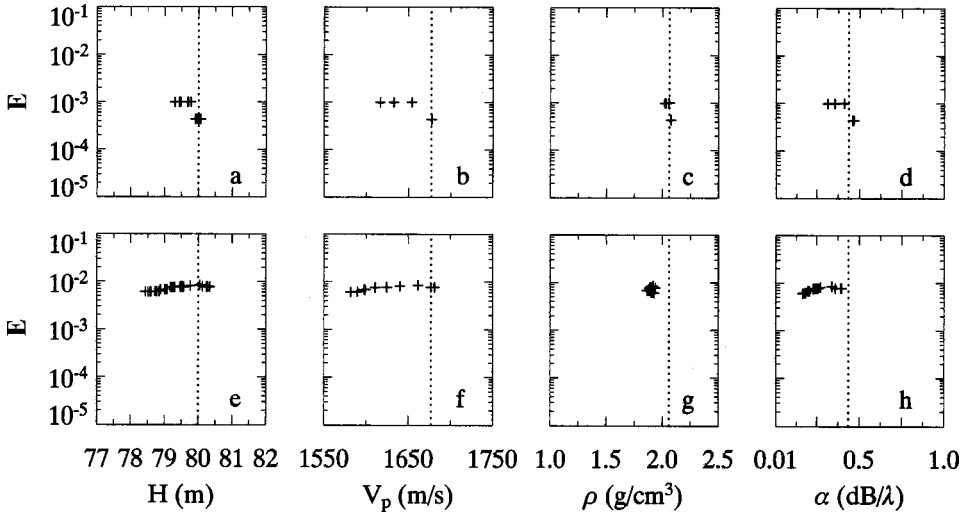


FIG. 5. Mismatch vs model parameter values for inversion solutions for studies V2(a) shown in (a)–(d) and V2(b) shown in (e)–(h) are presented ($s_{vx} = \pm 20$ m/s). For each study, a unique range-dependent environment was used. The model parameters, abscissa limits, and vertical dotted lines are as described in Fig. 3.

of noise to synthetic noise-free acoustic data, which can reduce the precision at which model parameters can be determined.[1,28]

The effect of environmental variance of sediment sound speed is striking. The value $|s_{vx}| = 40$ m/s is low compared to reported natural variations in surficial sound speed in shallow water. McKay *et al.*[18] reported standard deviations of sound speed of 71 m/s over hundreds of kilometers on the Scotian shelf, while Mayer *et al.*[29] measured a sound speed range of 277 m/s in the ONR Geoclutter area, even measuring a difference of 118 m/s between stations less than 1 km apart. (No standard deviations were published.) Lepage[30] found strong effects in propagation and reverberation model results from an input sound speed variation of 20 m/s. The results in this paper show that inversions in environments with typical values of sound speed variance lead to significant uncertainty in all estimated parameter values, when environmental variability is ignored.

It is also informative to look at the correlations between parameters. Figures 7–9 show two-dimensional correlation scatter plots for studies V1, V2, and V3, respectively. As seen earlier, in study V1 for $s_{vx} = \pm 5$ m/s there are groups of results about m_{vx} and the true model parameter values but there are also other solutions away from those values. Those other solutions typically have lower values than m_{vx} or the true parameter values. It can also be seen that the results are highly correlated between parameters. In Fig. 8, the V2 study data for $s_{vx} = \pm 20$ m/s are presented. The parameters are no longer as highly correlated with the exception of V_p and H. The spreads of the results have increased but there are still solutions clustered about m_{vx} and the true parameter values, as seen earlier in Figs. 6(e)–6(h). Like for the V1 study, the inversion solutions for V_p and H are not likely to be greater than m_{vx} and the true H parameter value. In Fig. 9, the V3 study data for $s_{vx} = \pm 40$ m/s are presented. In this figure, the distributions of the inversion results have increased again. There is no cluster about m_{vx} or the true α parameter value. However, V_p and H are still highly correlated. The strong correlations between inverted parameters suggest that for moderate data/model mismatch, there is an ambiguity between members of a family of models, where each model in

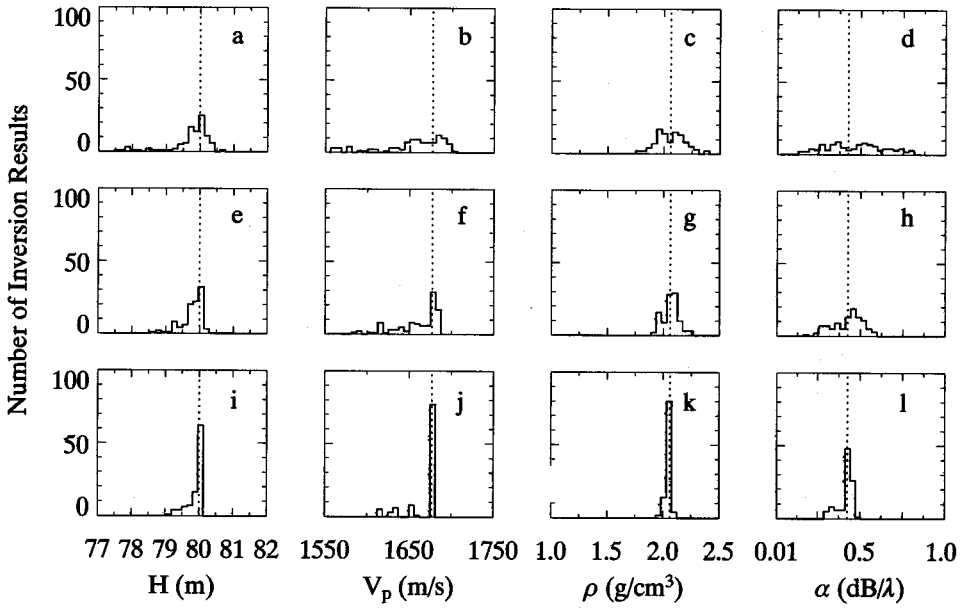


FIG. 6. Comparing different s_{vx} through histograms. The histograms of the inversion results for studies V1 (i)–(l), V2 (e)–(h), and V3 (a)–(d) are presented. The inversion results include model parameters values for water depth H, compressional speed V_p, density ρ, and attenuation α. For H, ρ, and α the vertical dotted lines represent the true model parameter values used to generate the measured acoustic field, and for V_p it represents the m_{vx}. The abscissa limits represent the bounds used in the inversions.

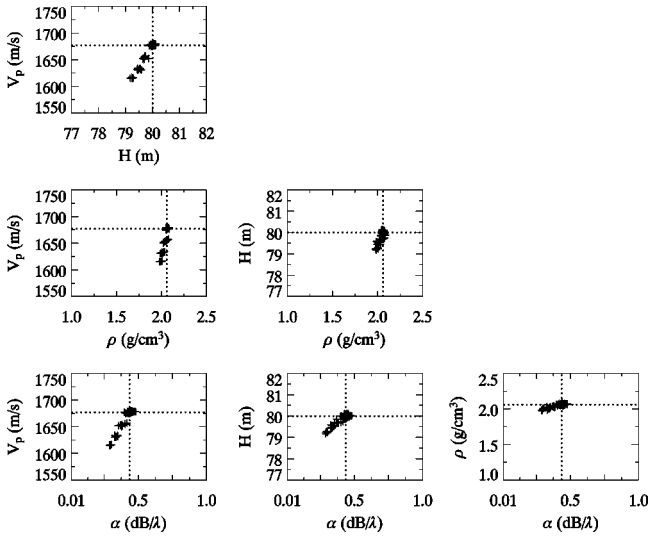


FIG. 7. The two-dimensional correlation scatter plots for the inversion solutions of study V1. The inversion results include model parameters values for H, V_p, ρ, and α. For H, ρ, and α the dotted lines represent the true model parameter values used to generate the measured acoustic field, and for V_p it represents the m_{vx}.

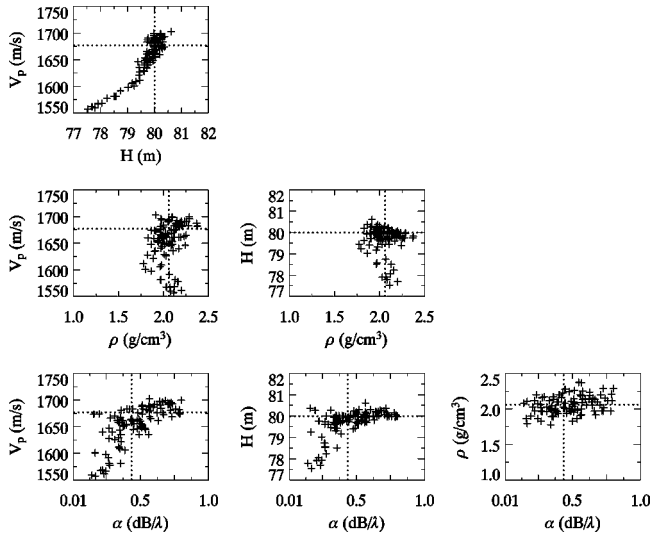


FIG. 9. The two-dimensional correlation scatter plots for the inversion solutions of study V3. The inversion results include model parameters values for H, V_p, ρ, and α. The dotted lines are as described in Fig. 7.

a family has a similar net effect on the acoustic field. In other words, the inversion has found a family of ''effective'' or nearly ''equivalent'' seabed models, as described by Chapman.[27] As the environmental variability increases, the data/model mismatch becomes greater, and the correlation worsens. Again, for water depth and compressional speed, parameter values above certain values are still not accepted.

To examine the effect of different values of R on the inversion results, a study similar to V2 was done using a range of 15 km (study V4) rather than 4 km. The standard deviation of the variability was $s_{vx}=\pm 20$ m/s. The results of study V4 are shown in Figs. 10(a)–10(d) along with the comparable V2 results in Figs. 10(e)–10(h). Between these ranges, the overall differences are few. The values of H and V_p have higher variances at the longer range and are skewed

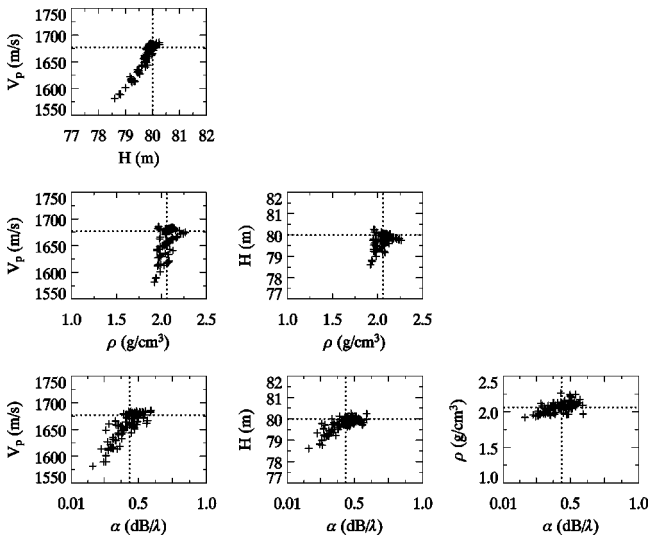


FIG. 8. The two-dimensional correlation scatter plots for the inversion solutions of study V2. The inversion results include model parameters values for H, V_p, ρ, and α. The dotted lines are as described in Fig. 7.

above the true mean values. This implies that the acoustic field is less complex at longer range, providing less independent information across the array of sensors. Another possibility is that the effects of the variable environment accumulate with range, making the data fields more noise-like, which can make precisely determining the true and true mean parameter values more difficult.

Similar studies to V2 were also performed using different source frequencies F. The results are shown in Fig. 11 through histogram plots. Figure 11 presents data for studies V5, V6, V2, and V7 (Table II) which have source frequencies $F=25$, 50, 100, and 200 Hz, respectively. For these studies $R=4$ km and $s_{vx}=\pm 20$ m/s. It should be noted that for the 200 Hz study (V7) the thickness of the false bottom was changed during the study. It was determined that a 10 m absorbing layer thickness (ΔH_a) was equivalent to a 100 m thickness. This change decreased the computational time of each call to PECan. The inversion results for the 25 and 50 Hz data [Figs. 11(m)–11(p) and 11(i)–11(l), respectively] show no preference for any particular model parameter values. This is probably due to the lack of information available in the acoustic fields. Drawing from normal mode theory, the 25 Hz source has one trapped mode traveling through the environment. The 50 Hz source has approximately three trapped modes. Due to the simplicity of the acoustic fields, it is perhaps logical that the fields do not have enough information to help the inversion algorithm determine the appropriate model parameters for this environment. For 200 Hz source data [Figs. 11(a)–11(d)], the clustering about m_{vx}, and the true parameter values of H and ρ are more pronounced than the 100 Hz results [Figs. 11(e)–11(h)]. These results imply that the complexity of the acoustic field increases with frequency, at least for this simple single-layer model. Consequently, at the higher frequencies, the data at different sensors would be more independent, while at low frequencies the data would be more correlated across the

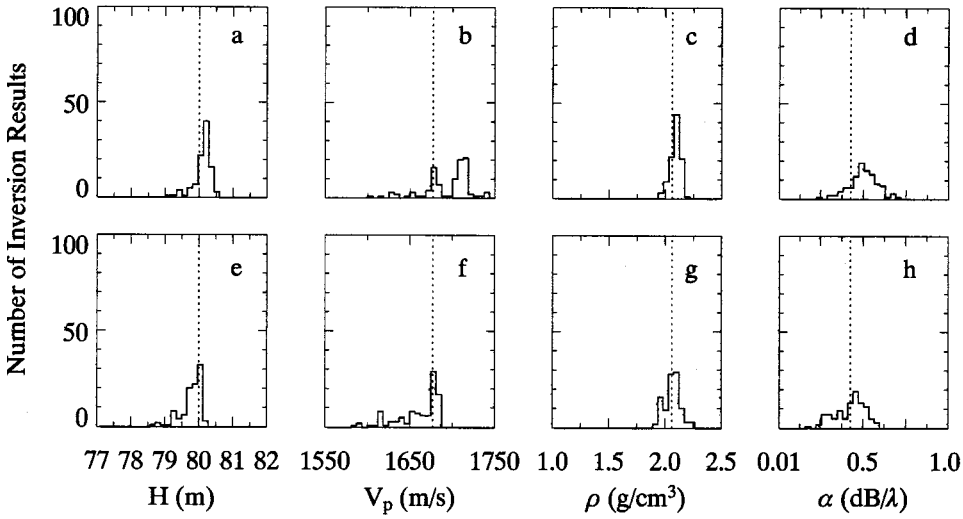


FIG. 10. Comparing different ranges. The histograms of the inversion results for studies V4 ($R=15$ km) shown in (a)–(d) and V2 ($R=4$ km) shown in (e)–(h) are presented. The inversion results include model parameters values for H, V_p, ρ, and α. The vertical lines and abscissa limits are as described in Fig. 6.

array. In Fig. 11, even with ignored variability, the uncertainty in the model parameters seems to decrease at higher frequencies; however, this may not be the case for multilayer environments, in which the high frequencies naturally would be more sensitive to the parameters of the upper layers than those of the deeper layers.[31–33] The effects of ignored variability on multilayer inversion are proposed as a topic of further study, along with the potential remedy of multifrequency inversion.

Finally, with respect to variable compressional speed, a multiple source frequency study was completed (study V8), i.e., a study in which "measured" acoustic data from more than one source frequency was used at the same time to do the inversion. The source frequencies of 200 and 100 Hz were used. Figure 12 shows the histogram results of study V8 with other comparable single source frequency results. The single frequency results shown in Fig. 12 are for $F=100$ Hz [Figs. 12(i)–12(l)] and for $F=200$ Hz [Figs. 12(e)–12(h)]. It can be seen, in comparison to the single frequency results, that the algorithm was more successful at converging to the true parameter values for ρ and H when the multiple frequencies were used. Approximately 90% of the predicted H parameter values were very near the true value. This encouraging result suggests that multifrequency inversion provides a partial remedy for uncertainty in inversion parameters in problems with ignored seabed variability, a possible topic for future work.

Table III states the mean of the inversion solution parameter values for the variable properties for each of the studies. As can be seen from Table III, the true mean of the studies, m_{vx}, always fell within one standard deviation of the mean of the inversion solution values for each study. However, there was considerable difference from study to study between the values of s_{NV}, the standard deviation of the model parameter values for compressional speed, as would be expected from the results. It should be noted that care should be taken in comparing s_{NV} to s_{vx} (Table II). The values of s_{vx} are the standard deviations about m_{vx} which was used to generate range-dependent variability while s_{NV} are standard deviations of the respective samples of solutions for compressional speed about the sample mean.

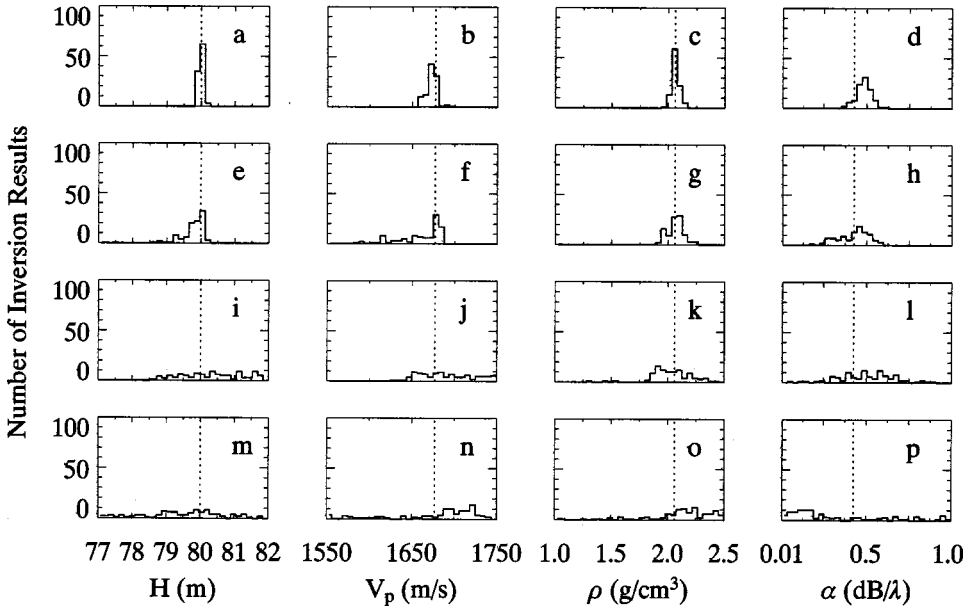


FIG. 11. Comparing different frequencies. The histograms of the inversion results for studies V7 ($F=200$ Hz) shown in (a)–(d), V2 ($F=100$ Hz) shown in (e)–(h), V6 ($F=50$ Hz) shown in (i)–(m), and V5 ($F=25$ Hz) shown in (n)–(q) are presented. The inversion results include model parameters values for H, V_p, ρ, and α. The abscissa limits represent the bounds used in the inversions. The vertical lines and abscissa limits are as described in Fig. 6.

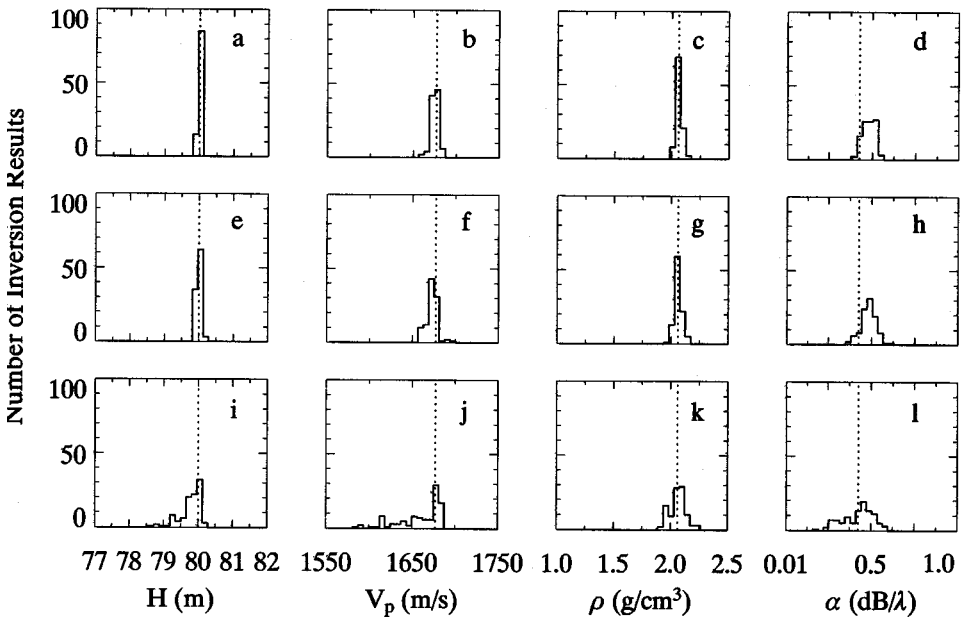


FIG. 12. Looking at multiple frequency inversion. The histograms of the inversion results for studies V8 ($F=100$ and 200 Hz) shown in (a)–(d), V7 ($F=200$ Hz) shown in (e)–(h), and V2 ($F=100$ Hz) shown in (i)–(l) are presented. The inversion results include model parameters values for H, V_p, ρ, and α. The vertical lines and abscissa limits are as described in Fig. 6.

B. Variable water depth

Similar to the studies done for variable compressional speed, studies were also done for variable water depth $h(x)$. The features of the studies H1–H5 referred to in this work for variable water depth are summarized in Table II. The source frequency used was 100 Hz and the range between source and receiver was 4 km. Like for variable compressional speed, the mismatches were calculated using Eq. (3), with $nF=1$ and $nS=1$. For studies H1–H5, 100 inversions were completed for each study, each inversion using a unique realization of the ocean bottom.

To generate the variable water depths, a combination of sine waves with random phases was used to create smooth and realistic variability. See Appendix B for the details of how the variability was created. For $N+1$ environmental profiles defined in PECan, $N+1$ water depth points were generated using the method described in Appendix B to form a set of values $\{[h]_i, i=0,N\}$. These values were assigned to the $N+1$ environmental profiles and PECan was allowed to linearly interpolate between the profiles to create $h(x)$. For studies H1–H5, the mean of the variability $m_{hx}=80$ m and the standard deviation of the variability s_{hx} were set as constants and defined as

$$m_{hx}=\frac{1}{N+1}\sum_{i=0}^{N}[h]_i, \tag{7}$$

$$s_{hx}=\pm\sqrt{\frac{1}{N+1}\sum_{i=0}^{N}([h]_i-m_{hx})^2}. \tag{8}$$

The s_{hx} of Eq. (8) is equivalent to s_{hx} of Eq. (B9). The distance between environmental profiles in PECan was 10 m.

On the left-hand side of Fig. 13, three realizations of the ocean bottom are shown, one realization from study H1 where $s_{hx}=\pm0.125$ m, one realization from study H3 where $s_{hx}=\pm0.5$ m, and one realization from H5 where $s_{hx}=\pm1$ m. It can be seen that for $s_{hx}=\pm0.125$ m the water depth barely deviates from the mean m_{hx}. The realizations of the ocean bottoms, such as those shown in Fig. 13, are a series of contiguous line segments, as mentioned earlier. As seen in the figure, the segments were small enough with respect to the entire range R that it made the realization of the ocean bottom appear to be smooth. On the right-hand side of Fig. 13 a histogram is shown of the slopes for each segment used to make each of the 100 realizations of the ocean bottoms for each of the three studies. For $s_{hx}=\pm0.125$ m, most segments have a slope within $\pm0.5°$. For $s_{hx}=\pm0.5$ m, most segments have a slope within $\pm1°$. For $s_{hx}=\pm1$ m, most segments have a slope within $\pm2°$. The distributions show no unexpected trends.

The two-dimensional correlation scatter plots for studies H1–H5 (not shown) had similar trends as for the variable compressional speed studies that allowed comparison be-

TABLE III. Sample mean (m) and sample standard deviation for the model parameter values of the compressional speed s_{NV} (studies V1–V8) and the water depth s_{NH} (studies H1–H5). The sample means can be compared to the set mean of the studies, $m_{vx}=1677$ m/s for V1–V8 and $m_{hx}=80$ m for H1–H5.

Study number	m (m/s) of V_p	s_{NV} (m/s)
V1	1668	±19
V2	1660	±26
V3	1655	±37
V4	1694	±28
V5	1680	±51
V6	1691	±32
V7	1672	±7
V8	1675	±4
Study number	**m (m) of H**	**s_{NH} (m)**
H1	79.8	±0.4
H2	79.7	±0.5
H3	79.6	±0.8
H4	79.5	±1.1
H5	79.6	±1.4

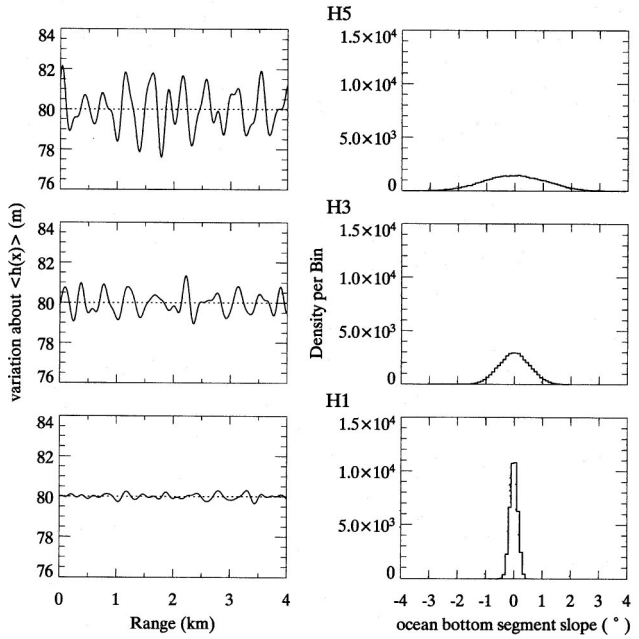


FIG. 13. Variable water depth. Sample realizations of the variable water depth are shown on the left-hand side for studies H1, H3, and H5. The dotted line represents the mean of the curve m_{hx}. On the right-hand side are histograms of all the slopes of the segments that make up each realization of the variable water depth used in the study.

tween different values of s_{vx} (Figs. 7–9). At $s_{hx}=\pm 0.125$ m, all parameters have two dimensional correlations. As s_{hx} increases the correlations decrease until almost no correlations are visible at $s_{hx}=\pm 1$ m, with the exception being a slight correlation between the V_p and H. The correlations that exist for low standard deviations again show the presence of a family of ''effective'' or nearly equivalent seabed models that can characterize the seabed.

Figure 14 shows the histograms for studies H1 [shown in Figs. 14(i)–14(l)], H3 [shown in Figs. 14(e)–14(h)], and H5 [shown in Figs. 14(a)–14(d)]. For $s_{hx}=\pm 0.125$ m [Figs. 14(i)–14(l)], there is a concentration of solutions about the true mean m_{hx} and true model parameter values but the concentration about these values decreases as s_{hx} increases. It appears that if one hopes to attain inversion results about the true or true mean values, then assuming a range-independent environment would be unwise for an s_{hx} value of 0.5 m or above, for this environment.

As was the case for sound speed, the effect of environmental variation of sound speed is striking. A shallow water seabed roughness of the order of 1 m is typical, as reported by Turgut *et al.*[34] for the New Jersey shelf. The results in this paper show that inversions in environments with typical values of water depth variance lead to significant uncertainty in all estimated parameter values, when environmental variability is ignored.

Table III reports the mean of the inversion solution parameter values for the variable water depth studies. Like for the variable compressional speed studies, the true mean of the studies, m_{hx}, always fell within one standard deviation of the mean of the inversion solution values for each study. And again, there was a difference from study to study between the values of s_{NH}, the standard deviation of the model parameter values for the water depth.

IV. CONCLUSIONS

Ignored variability in the environment can lead to significant uncertainty in the values of inversion parameters. Naively, one would hope that a single inversion of constant model parameters in a random environment would provide a confident estimate of the average environment, that is, the precise values of those parameters that do not vary and the mean values of those parameters that do vary. Typical experimental inversions of real ocean data implicitly make this assumption: it is impractical to model all the details of the random environment, so a simplified model with average parameters is devised, and the inversion is assumed to estimate those parameter values. This investigation has shown that matched-field geoacoustic inversion in variable environments is uncertain: the best-fit model of a single realization

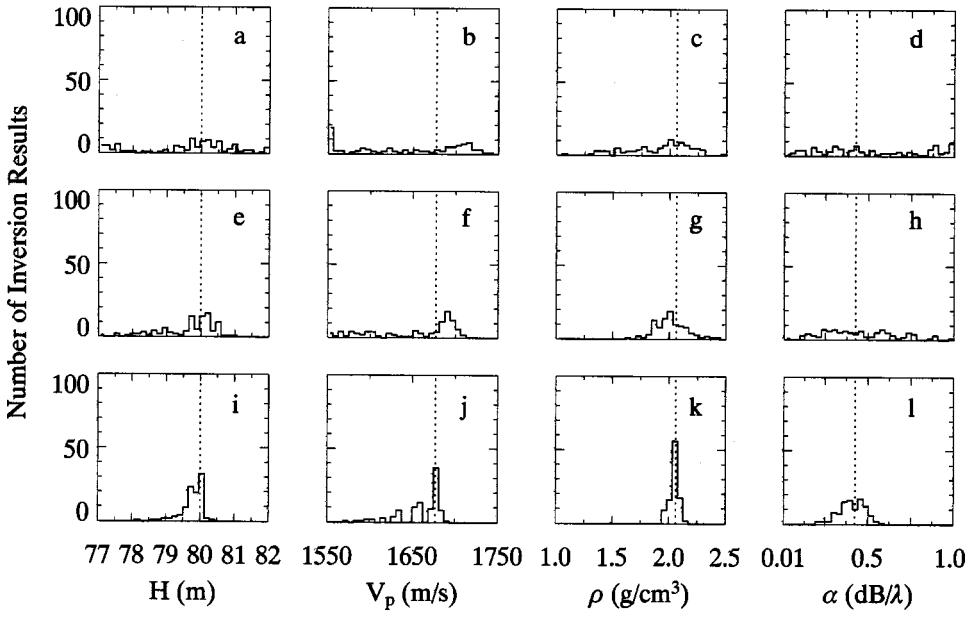


FIG. 14. Looking at the effects of variable water depth. The histograms of the inversion results are presented for studies H5 ($s_{hx}=\pm 1.0$ m) shown in (a)–(d), H3 ($s_{hx}=\pm 0.5$ m) shown in (e)–(h), and H1 ($s_{hx}=\pm 0.125$ m) shown in (i)–(l). The inversion results include model parameter values for water depth H, compressional speed V_p, density ρ, and attenuation α. For V_p, ρ, and α the vertical dotted lines represent the true model parameter values used to generate the measured acoustic field, and for H it represents the m_{hx}. The abscissa limits represent the bounds used in the inversions.

of the random environment does not necessarily reproduce the mean environment. The uncertainty is complex, that is, randomizing only one parameter in the environment leads to uncertainty in all parameters of the bestfit model. In addition, the uncertainty is structured, in that the variations of model parameters over all realizations of the random environment are correlated.

Looking at the statistics of the model parameter variations, it is clear (and no surprise) that the uncertainty of the inversions increases with the variability of the input model, regardless of which parameter is randomized. In many ways, the results are reminiscent of inversion of noisy acoustic data. The dependence of the inversion results on source-receiver range is weak, although there is some suggestion that there is more spread in inverted parameter values at longer ranges. Inversion performance improves with increasing frequency, likely due to there being more independent information at the hydrophone elements. Performance improves again when data from two frequencies are jointly inverted.

Looking at the pairwise scatter plots of inverted parameters, it is evident that the inverted parameters are strongly correlated, although the correlation weakens with increasing variability of the environment. This implies that the random-environment inversion problem is somewhat overdetermined by the chosen model parametrization; that is, the best-fit models form a family that represents an effective seabed model having simpler characteristics.

Although this investigation is not comprehensive, the results obtained indicate that a single inversion of a random environment would not confidently characterize the mean environment, although it would provide parameter values for a plausible model environment that matches the data within reasonable error bounds.

ACKNOWLEDGMENTS

The authors would like to thank Ron Kessel for his discussions, insights and inputs concerning this work.

APPENDIX A: CALCULATING MEAN AND STANDARD DEVIATION FOR $v(x)$

For variable compressional speed, to create $v_p(x)$, fluctuations were generated using scaled pseudorandom numbers picked from a uniform distribution. For $N+1$ fluctuations, $N+1$ pseudorandom numbers were generated and then linearly interpolated to form a line-form such as the one illustrated in Fig. 2. To keep control over the statistics of the line-forms the mean and standard deviation needed to be derived. This appendix outlines the derivation used.

Given a function $f(x)$, as illustrated in Fig. 15, defined on $[0,R]$, formed by connecting $N+1$ random numbers $\{y_i, i=0,N\}$ picked from a uniform distribution, then the global mean of $f(x)$ can be defined as

$$m_{fx}=\frac{1}{R}\int_0^R f(x)dx=\frac{1}{R}\sum_{i=0}^{N}\int_{x_i}^{x_{i+1}} f(x)dx, \tag{A1}$$

where $x_0=0$ and $x_{N+1}=R$. Let,

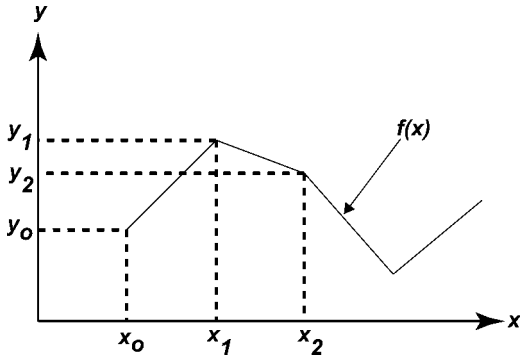


FIG. 15. Schematic representation of a line-form function [similar to $v_p(x)$] used to help clarify the derivation of the global mean and standard deviation of $f(x)$ in Appendix A.

$$f(x)=y=\left(\frac{y_{i+1}-y_i}{x_{i+1}-x_i}\right)(x-x_i)+y_i; \tag{A2}$$

therefore,

$$x=(y-y_i)\left(\frac{x_{i+1}-x_i}{y_{i+1}-y_i}\right)+x_i \tag{A3}$$

and

$$\frac{dx}{dy}=\left(\frac{x_{i+1}-x_i}{y_{i+1}-y_i}\right), \tag{A4}$$

$$dx=\left(\frac{x_{i+1}-x_i}{y_{i+1}-y_i}\right)dy. \tag{A5}$$

Given that $f(x_i)=y_i$ then substituting Eq. (A5) into Eq. (A1) and substituting y for $f(x)$ gives

$$m_{fx}=\frac{1}{R}\sum_{i=0}^{N}\int_{y_i}^{y_{i+1}} y\left(\frac{x_{i+1}-x_i}{y_{i+1}-y_i}\right)dy. \tag{A6}$$

If $x_{i+1}-x_i=\Delta x$, as is the case in this work, and given that $\Delta x/R=1/N$ then Eq. (A6) solves to be

$$m_{fx}=\frac{\Delta x}{R}\sum_{i=0}^{N}\frac{y_{i+1}^2-y_i^2}{2(y_{i+1}-y_i)}=\frac{1}{2N}\sum_{i=0}^{N}(y_{i+1}+y_i). \tag{A7}$$

Equation (A7) is therefore the global mean for the line-form $f(x)$.

Similarly, the square of the standard deviation s_{fx} for the line-form $f(x)$ can be derived thus,

$$\begin{aligned}
s_{fx}^2&=\frac{1}{R}\int_0^R (f(x)-m_{fx})^2dx\\
&=\frac{1}{R}\sum_{i=0}^{N}\int_{x_i}^{x_{i+1}}(f(x)-m_{fx})^2dx\\
&=\frac{1}{R}\sum_{i=0}^{N}\int_{y_i}^{y_{i+1}}(y-m_{fx})^2\frac{x_{i+1}-x_i}{(y_{i+1}-y_i)}dy\\
&=-m_{fx}^2+\frac{\Delta x}{3R}\sum_{i=0}^{N}\frac{(y_{i+1}^3-y_i^3)}{(y_{i+1}-y_i)}\\
&=-m_{fx}^2+\frac{1}{3N}\sum_{i=0}^{N}(y_{i+1}^2+y_i^2+y_{i+1}y_i).
\end{aligned} \tag{A8}$$

APPENDIX B: VARIABLE WATER DEPTH

For modeling purposes, the seabed height is simulated as a random process based on observed power-law wavenumber spectra. Let

$$\hat{h}(x)=\int_{-\infty}^{+\infty} S(k)e^{ikx}dk, \tag{B1}$$

in which $\hat{h}(x)$ is the height of the seabed above the mean level [i.e., $\hat{h}(x)+m_{hx}=h(x)$], x is the distance along the track, k is the wave number ($k=2\pi/\lambda$), λ is the wavelength, and $S(k)$ is the Fourier transform of $\hat{h}(x)$.

Since $\hat{h}(x)$ is real, it is necessary that

$$S(-k)=S^*(k), \tag{B2}$$

where * indicates the complex conjugate. It is convenient to represent $S(k)$ in polar form, that is,

$$S(k)=|S(k)|e^{i\phi(k)}, \tag{B3}$$

in which case the phase function $\phi(k)$ must have the symmetry

$$\phi(-k)=-\phi(k), \tag{B4}$$

so that Eq. (B2) is satisfied. Then Eq. (B1) becomes

$$\hat{h}(x)=2\int_0^{+\infty}|S(k)|\cos[kx+\phi(k)]dk. \tag{B5}$$

The covariance of the surface is

$$C(x)=\int_{-\infty}^{+\infty}\hat{h}(x')\hat{h}(x'+x)dx'$$

$$=2\int_0^{+\infty}|S(k)|^2\cos(kx)dk. \tag{B6}$$

Note that the covariance is independent of the phase function, and that the spectrum $|S(k)|^2$ is the Fourier transform of the covariance $C(x)$. By randomly choosing different phase functions, it is possible to create multiple realizations of a rough surface having uniform covariance and spectrum.

The square of the standard deviation of the surface height is simply

$$\sigma^2=C(0)=2\int_0^{+\infty}|S(k)|^2dk, \tag{B7}$$

that is, the area under the spectrum.

To implement this in simulations, we employ a finite sum over sinusoids having discrete wave numbers $k_n \in \{k_1,k_2,k_3,\ldots,k_N\}$, specified amplitudes $A_n \in \{A_1,A_2,A_3,\ldots,A_N\}$, and random phases $\phi_n \in \{\phi_1,\phi_2,\phi_3,\ldots,\phi_N\}$. That is,

$$\hat{h}(x)\approx\sum_{n=1}^{N}A_n\cos(k_nx+\phi_n). \tag{B8}$$

We need to choose the wave numbers, amplitudes, and phases to generate a seabed with the correct statistics.

The square of the standard deviation from Eq. (B8) is

$$s_{hx}^2\approx\frac{1}{2}\sum_{n=1}^{N}A_n^2\approx\sigma^2 \tag{B9}$$

so it seems that A_n^2 plays the same role as $|S(k)|^2dk$.

Observations have shown[35,36] that seabed roughness spectra can be described in a power-law form:

$$|S(k)|^2\propto k^{-b}(b>1), \tag{B10}$$

such that long wavelengths (low wave numbers) have higher amplitudes than short wavelengths (high wave numbers). In order that the integral in Eq. (B7) and the sum in Eq. (B9) be finite. It is necessary that the band of wave numbers be limited, at least at the lower end; that is,

$$k_{\min}<k<k_{\max}. \tag{B11}$$

Since A_n^2 plays the same role as $|S(k)|^2dk$, we propose that the amplitudes be given by

$$A_n=\sqrt{ak_n^{-b}\Delta k_n}, \tag{B12}$$

in which

$$a\approx\frac{2s_{hx}^2}{\Sigma_{i=1}^{N}k_i^{-b}\Delta k_i} \tag{B13}$$

and

$$\Delta k_n=\tfrac{1}{2}(k_{n+1}-k_{n-1}). \tag{B14}$$

It remains to choose the actual distribution of wave numbers. We propose that the wave numbers be distributed unevenly, progressively choosing more at the lower range, to minimize the number needed. One approach is to define the sequence

$$k_n=\alpha^{n-1}k_1, \tag{B15}$$

in which

$$k_1\equiv k_{\min}\equiv 2\pi/\lambda_{\max}, \tag{B16}$$

$$k_N\equiv k_{\max}, \tag{B17}$$

and α is a factor to be determined. Let k_N be the wave number at which the spectrum $|S(k)|^2$ is some fraction β of its value at k_1, that is,

$$\beta=\left(\frac{k_1}{k_N}\right)^b \tag{B18}$$

then, from Eq. (B15),

$$\alpha=\beta^{-1/[b(N-1)]}. \tag{B19}$$

Here is the procedure to synthesize a rough seabed: choose a standard deviation of the surface height s_{hx} (say, 0.5 m), a power-law exponent b (2–3 is typical), a maximum wavelength $\lambda_{\max}$, the number of sinusoidal components N (say, 10), and a value for β (say, 1/20). Then calculate α from Eq. (B19), k_1 from Eq. (B16), and the other k_n from Eq. (B15). From the wave numbers, calculate the Δk_n from Eq. (B14), and then the A_n from Eqs. (B12) and (B13). Then choose a set of phases $\{\phi_1,\phi_2,\phi_3,\ldots,\phi_N\}$ randomly distributed over an interval of 2π radians. Finally, calculate the random seabed function $\hat{h}(x)$ from Eq. (B8). Different sets of random phases will generate different realizations of the seabed having the same statistics. It is prudent to calculate

s_{hx} from the final result to verify that it has the value one started with. The method of synthesizing roughness profiles using discrete sine waves was adapted from the synthesis of ocean surface wave forms.[37] The parameter values used in this study are: $\lambda_{max}=500\,\mathrm{m}$, $N=9$, $b=3$, and $\beta=0.05$. This combination of values generates plausible synthetic realizations of seabed topography with reproducible statistics.

[1] S. E. Dosso, M. J. Wilmut, and A. S. Lapinski, ''An adaptive hybrid algorithm for geoacoustic inversion,'' IEEE J. Ocean. Eng. **26**, 324–336 (2001).

[2] M. R. Fallat and S. E. Dosso, ''Geoacoustic inversion via local, global and hybrid algorithms,'' J. Acoust. Soc. Am. **105**, 3219–3230 (1999).

[3] P. Gerstoft, ''Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach,'' J. Acoust. Soc. Am. **97**, 2181–2190 (1995).

[4] M. D. Collins, W. A. Kuperman, and H. Schmidt, ''Nonlinear inversion for ocean-bottom properties,'' J. Acoust. Soc. Am. **92**, 2770–2783 (1992).

[5] G. J. Heard, D. Hannay, and S. Carr, ''Genetic algorithm inversion of the 1997 Geoacoustic Inversion Workshop test case data,'' J. Comput. Acoust. **6**, 61–71 (1998).

[6] P. Ratilal, P. Gerstoft, and J. T. Goh, ''Subspace approach to inversion by genetic algorithms involving multiple frequencies,'' J. Comput. Acoust. **6**, 99–115 (1998).

[7] S. D. Rajan, J. F. Lynch, and G. V. Frisk, ''Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water,'' J. Acoust. Soc. Am. **82**, 998–1017 (1987).

[8] A. Tolstoy, ''MFP benchmark inversions via the RIGS method,'' J. Comput. Acoust. **6**, 185–203 (1998).

[9] N. R. Chapman, S. Chin-Bing, D. King, and R. B. Evans, ''Benchmarking geoacoustic inversion methods for range-dependent waveguides,'' IEEE J. Ocean. Eng. **28**, 320–330 (2003).

[10] G. H. Brooke, D. J. Thomson, and G. R. Ebbeson, ''PECan: A Canadian parabolic equation model for underwater sound propagation,'' J. Comput. Acoust. **9**, 69–100 (2001).

[11] A. S. Lapinski and S. E. Dosso, ''Bayesian geoacoustic inversion for the Inversion Techniques 2001 Workshop,'' IEEE J. Ocean. Eng. **28**, 380–393 (2003).

[12] P. L. Nielsen, M. Siderius, and P. Gerstoft, ''Range-dependent geoacoustic inversion: Results from the inversion techniques workshop,'' IEEE J. Ocean. Eng. **28**, 414–423 (2003).

[13] C. H. Harrison and M. Siderius, ''Effective parameters for matched field geoacoustic inversion in range-dependent environments,'' IEEE J. Ocean. Eng. **28**, 432–445 (2003).

[14] M. Siderius, P. L. Nielsen, and P. Gerstoft, ''Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array,'' J. Acoust. Soc. Am. **112**, 1523–1535 (2002).

[15] M. R. Fallat, S. E. Dosso, and P. L. Nielsen, ''An investigation of algorithm-induced variability in geoacoustic inversion,'' IEEE J. Ocean. Eng. **29**, 78–87 (2004).

[16] D. M. F. Chapman and D. D. Ellis, ''Propagation loss modelling on the Scotian Shelf: the geo-acoustic model,'' in *Bottom-Interacting Ocean Acoustics*, edited by W. A. Kuperman and F. B. Jensen (Plenum, New York, 1980); the proceedings of a conference held at SACLANTCEN, La Spezia, Italy, June 1980.

[17] D. D. Ellis and D. M. F. Chapman, ''Propagation loss modelling on the Scotian Shelf: comparison of model predictions with experiment,'' in *Bottom-Interacting Ocean Acoustics*, W. A. Kuperman and F. B. Jensen, eds. (Plenum Press, New York, 1980); the proceedings of a conference held at SACLANTCEN, La Spezia, Italy, June 1980.

[18] A. G. McKay, J. A. Hunter, R. L. Good, and D. M. F. Chapman, ''A 12-channel marine EEL for shallow refraction surveying of the seabottom in coastal waters,'' in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berson (Plenum, New York, 1986), pp. 871–880.

[19] J. C. Osler, ''A geo-acoustic and oceanographic description of several shallow water experimental sites on the scotian shelf,'' DREA Technical Memorandum 94/216 [Defence Research Establishment Atlantic (now Defence Research Development Canada—Atlantic), Dartmouth, Nova Scotia, 1994].

[20] E.-L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

[21] S. E. Dosso and M. J. Wilmut, ''Effects of incoherent and coherent source spectral information in geoacoustic inversion,'' J. Acoust. Soc. Am. **112**, 1390–1398 (2002).

[22] C. F. Mecklenbrauker and P. Gerstoft, ''Objective functions for ocean acoustic inversion derived by likelihood methods,'' J. Comput. Acoust. **8**, 259–270 (2000).

[23] J. A. Nelder and R. Mead, ''A simplex method for function minimization,'' Comput. J. **7**, 308–313 (1965).

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[25] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, ''Optimization by simulated annealing,'' Science **220**, 671–680 (1983).

[26] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, ''Equation of state calculation by fast computing machines,'' J. Chem. Phys. **21**, 1087–1092 (1953).

[27] D. M. F. Chapman, ''What are we inverting for?,'' in *Inverse Problems in Underwater Acoustics*, edited by M. I. Taroudakis and G. N. Makrakis (Springer, New York, 2001), pp. 1–14.

[28] S. E. Dosso, ''Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach,'' J. Acoust. Soc. Am. **111**, 129–142 (2002).

[29] L. A. Mayer, B. J. Kraft, P. Simpkin, P. Lavoie, E. Jabs, and E. Lynskey, ''In-situ determination of the variability of seafloor acoustic properies: an example form the ONR geoclutter area,'' in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, Dordrecht, 2002), pp. 115–122.

[30] K. D. Lepage, ''Modeling propagation and reverberation sensitivity to oceanographic and seabed variability,'' in Ref. 29, pp. 353–360.

[31] M. D. Collins, W. A. Kuperman, and H. Schmidt, ''Nonlinear inversion for ocean-bottom properties,'' J. Acoust. Soc. Am. **92**, 2770–2783 (1992).

[32] P. Ratilal, P. Gerstoft, and J. T. Goh, ''Subspace approach to inversion by genetic algorithms involving multiple frequencies,'' J. Comput. Acoust. **6**, 99–115 (1998).

[33] M. I. Taroudakis and M. G. Markaki, ''Bottom geoacoustic inversion by matched field processing—A sensitivity study,'' Inverse Probl. **16**, 1679–1692 (2000).

[34] A. Turgut, D. Lavoie, D. J. Walter, and W. B. Sawyer, ''Measurements of bottom variability during SWAT New Jersey shelf experiments,'' in Ref. 29, pp. 91–98.

[35] J. M. Berkson and J. E. Mathews, ''Statistical properties of sea-floor roughness,'' in *Acoustics and the Seabed*, edited by N. G. Pace (Bath University Press, Bath, UK, 1983), pp. 215–223.

[36] K. B. Briggs, D. Tang, and K. L. Williams, ''Characterization of interface roughness of rippled sand off Fort Walton Beach, Florida,'' IEEE J. Ocean. Eng. **27**, 505–514 (2002).

[37] J. P. Brett, ''Computer-generated random sea surfaces for use with the program CABUOY,'' Naval Air Development Center Technical Memorandum 2063-TM-46-77, July, 1977.

Model-oriented ocean tomography using higher frequency, bottom-mounted hydrophones

James K. Lewis
Scientific Solutions, Inc., 4875 Kikala Road, Kalaheo, Hawaii 96741

Jason Rudzinsky
Applied Physical Sciences, Corp., 2 State Street, Suite 300, New London, Connecticut 06320

Subramaniam Rajan, Peter J. Stein, and Amy Vandiver
Scientific Solutions, Inc., 99 Perimeter Road, Nashua, New Hampshire 03063

The KauaiEx Group

(Received 1 November 2004; revised 22 February 2005; accepted 23 February 2005)

A tomographic scheme is presented that ingests ocean acoustic measurements into an ocean model using data from bottom-mounted hydrophones. The short distances between source-receiver pairs (1–10 km) means arrival times at frequencies of 8–11 kHz are readily detectable and often distinguishable. The influence of ocean surface motion causes considerable variability in acoustic travel times. Techniques are presented for measuring travel times and removing the variability due to surface waves. An assimilation technique is investigated that uses differences in measured and modeled acoustic travel times to impose corrections on the oceanographic model. Equations relating travel time differences to oceanographic variables are derived, and techniques are presented for estimating the acoustic and ocean model error covariance matrices. One test case using a single source-receiver pair shows that the tomographic information can have an impact on constraining the solution of the ocean circulation model but can also introduce biases in the predictions. A second test case utilizes knowledge of a bias in a model-predicted variable to limit grid cells that are impacted by the tomographic data. In this case, using the tomographic data results in significant improvements in the model predictions without introducing any biases.
[DOI: 10.1121/1.1893355]

PACS numbers: 43.30.Pc [AIT] Pages: 3539–3554

I. INTRODUCTION

Variations in the travel times of acoustic signals are known to be related to the spatial and temporal changes in water column temperatures, salinities, and currents. As such, acoustic tomography can provide relevant and important information on the characteristics of the water column.[1] Moreover, the travel time of an acoustic pulse from fixed and well-separated source and receiver locations may be considered a more robust measure of spatially averaged oceanographic variables (e.g., sound speed or temperature) than are point measurements.[2,3] This results from the fact that the acoustic parameter is an integral over space, while point measurements are susceptible to local, small-spatial scale noise.

In the work presented here, we develop a tomographic formulation that uses travel time measurements of relatively high frequency (~10 kHz) acoustic transmissions across fixed, omnidirectional, bottom-mounted acoustic transducers. Measured travel times along distinct ray paths are expressed as the difference between a reference travel time (through an environment with a reference sound speed structure) and a weighted sum of sound speed perturbations (with respect to the reference sound speed structure) in a gridded plane containing the source and receiver. To constrain the solution of the sound speed perturbation vector, a spatial covariance structure is imposed. Instead of directly inverting for the sound speed perturbation (i.e., inverse tomography), we utilize a suboptimal assimilation process in which expressions relating the acoustic travel time measurements to ocean variables are used to ingest the acoustic data into a dynamic ocean model of the region. A spatial covariance matrix weights the acoustic observations and spatially distributes their influence throughout the ocean model domain. The assimilation scheme allows the acoustic observation to locally influence both nowcasts and forecasts of variables in the ocean model (e.g., water temperature). Our system is somewhat similar to that which would be implemented using a dynamic state space/parameter estimation scheme (e.g., a Kalman filtering approach[4]) except that dynamic updates to the covariance matrix are not performed with every new set of acoustic observations and model fields.

As a test bed for our tomographic studies, we have used two primary assets. The first is Pacific Missile Range Facility (PMRF) off the west coast of Kauai, Hawaii. PMRF has 15 bottom-mounted sources that operate in the 8–11-kHz band. PMRF has an additional 178 bottom-mounted receivers. These assets, used by PMRF primarily for localization, communications, and safety purposes, provide for the capability of transmitting and receiving acoustic data throughout the range. The distances between source-receiver pairs in the shallow regions of the range are relatively small (as short as 1 km). The arrival times in the 8–11-kHz band are readily detectable. Although the configuration of all hydrophones

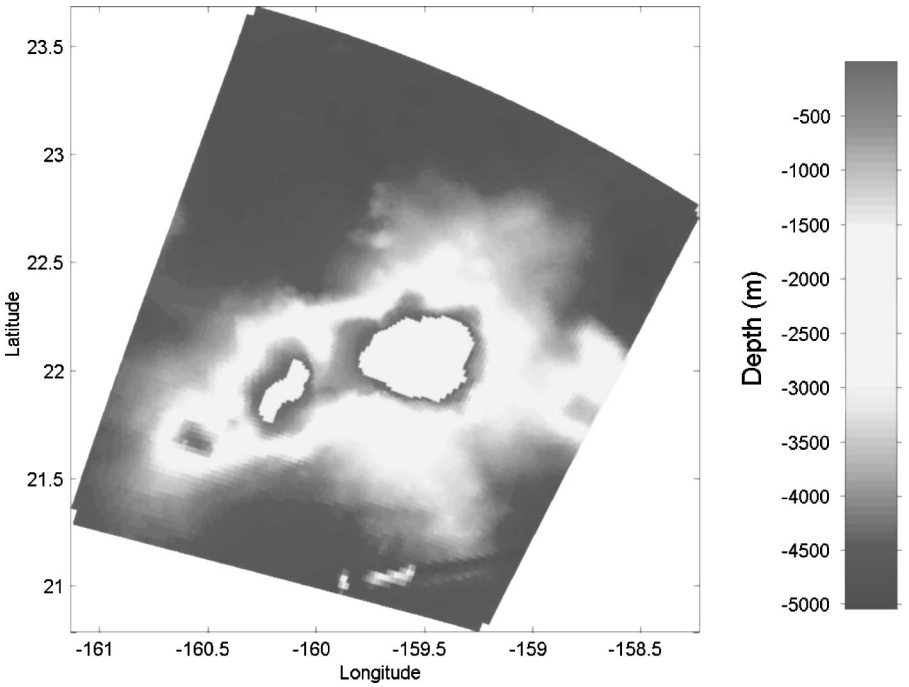


FIG. 1. Domain and depth field for the Kauai (island to the east) and Niihau (island toward the west) ocean model. PMRF reaches from between the islands northward to ~23 °N.

being bottom-mounted is unconventional, such an arrangement has tomographic benefits in that there is effectively no positional uncertainty of the sensors. However, the acoustic signals of greatest use from bottom-mounted systems interact with the ocean surface and are, therefore, subject to the influence of surface waves. As a result, arrival times in the relatively high-frequency range of interest (8–11 kHz) can have considerable short-term variability due to the surface wave fields, often the dominate source of variability.

The second asset we employed is the ocean circulation model that encompasses the waters of PMRF off the west end of the island of Kauai (Fig. 1). The authors developed this model for the purposes of obtaining estimates of the spatial and temporal variations of oceanographic parameters that would impact sound speed structure. The hydrodynamic model is executed on a daily basis, predicting temperatures (T), salinities (S), and currents ($\mathbf{U}$) at horizontal resolutions as high as 900 m and 28 levels in the vertical. The model provides a means of specifying realistic sound speed structure in space and time under varying tidal, atmospheric, and wave conditions. As such, the model offers a means of estimating acoustic ray arrival times in the ocean surrounding PMRF. In addition, the ocean model provides the testing capability of a method for assimilating acoustic travel times to impact model-calculated T, S, and $\mathbf{U}$ using a formulation that is an analog of the physical-space statistical analysis system (PSAS) data assimilation scheme.[5]

Using our model-oriented assimilation approach (directly relating acoustic data to the T, S, and $\mathbf{U}$ of a model) has some distinct benefits for providing more accurate model predictions. First, acoustic data reflect information about a volume within the water column and provide a natural means for obtaining spatially averaged measures of oceanographic variables. This is opposed to single point or vertical profile observations that are susceptible to small spatial-scale variations, noise, and uncertainties that may limit their utility as ocean model inputs. Second, the PSAS scheme allows even just one source-receiver pair to influence the model solutions in a large volume of water surrounding the path of the acoustic ray between the source and receiver. This is achieved utilizing estimates of the spatial error covariance matrices for the acoustic observations and the model variables.

Here we present formulations for quantifying the observed arrival times of acoustic rays whose paths have interacted with the ocean surface. Arrival time anomalies are determined relative to a monthly sound velocity structure based on the three-dimensional grid structure of the hydrodynamic model of the PMRF region. We present transforms for determining a model-related travel time anomaly along the path of the transmitted acoustic ray. It is shown that the differences between the model and observed travel time anomalies can be transformed back to adjustments of model-predicted water temperatures, salinities, and currents.

There are several elements involved with this work. These include collection and analysis of arrival times and arrival time anomalies for specific acoustic paths, the details of the numerical ocean model, and the specific construct of the PSAS to relate travel time anomalies to model-predicted variables for assimilation into the ocean model. Each element will be discussed as well as the details of test cases in which travel times from one source-receiver pair were assimilated into the ocean model.

II. THE OCEAN MODELS FOR PMRF

A hydrodynamic model, an adaptation of the Blumberg and Mellor model,[6] has been implemented for the waters surrounding Kauai and Niihau, Hawaii (Fig. 1). This particular version of the model uses a semi-implicit solution scheme for solving for the sea surface height field[7] and a hybrid z-level coordinate system in the vertical[8] to minimize problems that can arise with the original bottom-following sigma-

coordinate system when using steep bathymetry and realistic temperature and salinity profiles. The bathymetry (Fig. 1) is from the Smith-Sandwell topography[9] augmented with higher resolution sounding data obtained from NOAA and PMRF. A fundamental length scale that characterizes some of the flow-field activity in the region is the baroclinic radius of deformation, approximately 20 km. Therefore, the model grid spacing is set at 2–3 km around the open boundaries of the domains to approximately 1 km around the shoreline of Kauai. In addition, the total horizontal domain is well beyond 20 km off the islands.

Observed T-S characteristics were used in specifying the vertical resolution of the model. We chose the vertical grid structure (28 active levels) with higher resolution within the top 100 m and at those depths at which salinity extremes exist. This allows advective inflow conditions specified at the open boundaries to better maintain the observed T-S structure within the model domain.

We utilized the open boundary condition presented in Lewis *et al.*,[10] specifying the tidal sea level elevations and phases along the open boundaries of the model domains. The boundary values were obtained from the Oregon State University tidal model TPXO.3[11] but ''tuned'' to match observed amplitudes and phases for Kauai. The model was forced at the open boundaries with the M_2, S_2, N_2, O_1, K_1, and P_1 tidal constituents. The model error for the largest tidal constituent, the M_2 tide, is less than 4%, and the errors for the other constituents are similar or smaller in magnitude.

In addition to the ocean circulation, surface waves were modeled using Delft University of Technology's SWAN (simulating waves nearshore).[12,13] SWAN is a two-dimensional wave spectra model that can perform using a curvilinear-orthogonal grid. The two-dimensional spectra ability results in being able to realistically simulate the wide range of wave conditions typically encountered in the world's oceans. A curvilinear-orthogonal grid allows SWAN to use the same computational grid (and associated depths) as that used by the ocean circulation model. The means of allowing the interactions between currents and waves is greatly facilitated with the use of the same grids by both the wave and circulation models.

A. Initialization and forcing fields

In addition to tides, the model utilizes the Navy's daily modular ocean data assimilation system (MODAS) as a daily estimate of the three-dimensional *T*-*S* structure within the model domain. This is used to introduce the mesoscale circulation field into the model domain. MODAS fields use satellite sea surface temperatures, satellite altimetry, bathythermograph data, and results from other models to determine the *T*-*S* structure within a region. Our ocean model employs a scheme that nudges[14] the model-predicted temperatures and salinities to the MODAS temperatures and salinities. The nudging parameter T^* had a value of 0.75 days, representing a fairly strong nudge. However, tests showed that the model-predicted *T* and *S* often had substantial variations away from the MODAS *T*-*S* values as dictated by the governing physics within the model.

Atmospheric forcing from the National Centers for Environmental Prediction (NCEP) includes momentum, heat, and mass fluxes at the air-sea interface. With these, the ocean model can include wind forcing, precipitation minus evaporation, and sensible, evaporative, and radiative heat fluxes. The SWAN model is forced by the NCEP surface wind velocity and wave fields that are generated far from Hawaii and then propagate to the islands. To account for this latter factor, SWAN ingests NOAA's WaveWatchIII wave spectra information along the open boundaries of the model domain.

Since waves can have a significant impact on ocean circulation,[15] the ocean circulation model utilized the surface wave model results to calculate (1) wave-enhanced bottom friction, (2) Stokes drift and the Coriolis wave stress, (3) radiation stresses, (4) wave-related mixing length at the ocean surface, and (5) the virtual tangential surface stress.

B. Model accuracy

As a measure of the accuracy of the ocean circulation model, predictions and corresponding observations of water temperature were used to calculate an rms model error as a function of depth. During June and July 2003, 175 bathythermographs (BTs) were collected. Model temperatures were interpolated in space and time to these BT data to calculate the rms error for *T*. Values of rms errors range from $\Delta T = 0.22\,^{\circ}\mathrm{C}$ to $0.96\,^{\circ}\mathrm{C}$, which translates to sound speed errors of 1.2–4.8 m/s, assuming $\partial c/\partial T \approx 4.947$ m/s/°C, where *c* is sound speed. Since salinity does vary significantly with depth in the Hawaii region, errors in predicting water temperature will be the primary cause of errors in sound speed in the water column.

III. CONSTRUCTS RELATED TO THE PSAS DATA ASSIMILATION

Some of the forcing fields contain observations, such as the MODAS fields and the initial conditions of the atmospheric fields. But there is very little water column data represented within the forcing data, especially for forcing fields representing future conditions. Acoustic travel time information represents *in situ* data that could be assimilated into the ocean model to constrain the solutions of the model, and hopefully increase the accuracy of the model.

A. Simulating acoustic paths and travel times

The ocean model *T*-*S* structure can be used to calculate sound speed profiles (SSP) using expressions relating *T*, *S*, and depth to sound speed. This was done to obtain SSPs at a vertical resolution of 5 m and a horizontal resolution of 150 m. Knowing the precise locations of the PMRF sources and receivers, we used the SSP in the Bellhop acoustic propagation model to calculate paths of acoustic rays between the bottom-mounted sources and receivers as a function of time. We found that in many instances direct-path rays and single-surface bounce rays had arrival times very close to one another. This would have made it difficult to resolve these arrivals in field data due to finite source bandwidth. This was verified with actual field data, with the true environment often resulting in acoustic signatures that were even more complex than those predicted using model SSPs.

Our simulations using model results indicated that multiple surface bounce acoustic rays were stable in the path they took from source to receiver. Moreover, their arrival times were well separated from the early direct-path and single-surface bounce arrivals. As a result, their arrival times could be estimated fairly well. This provided a basis in our analysis of actual acoustic data for delineating times to consider for the arrival of rays traveling along specific paths. Moreover, multiple surface bounce rays provide a better sampling of the water column. Due to the drop in signal-to-noise for higher multiple surface bounce acoustic paths, we concentrated on analyzing paths that bounced off the ocean surface only twice.

Although acoustic simulation models are fairly accurate, they do have some limitations. For example, our modeling does not simulate the Doppler impacts of the motion of surface waves or surface roughness. However, simulations show that these factors can be effectively eliminated in the observations by averaging the arrival times over a number of pings that cover many cycles of surface motion. Another possible limitation is the error resulting from inaccuracies in the exact positions and depths of the source, receiver, and where a ray path bounces off the ocean bottom between the source and receiver.

Our analyses require a reference sound speed structure c_R. This was determined using monthly climatological *T*-*S* fields for the region shown in Fig. 1. Thus, c_R is a four-dimensional field with spatial resolution equal to that of the ocean circulation model and varying monthly. In addition, the PSAS assimilation scheme requires reference fields for *T*, *S*, and current velocities. Again, the monthly climatological *T*-*S* fields were used, while a reference velocity of 0 m/s was used throughout space and time.

B. Observed travel time anomalies

At a given time, an observed travel time t_o for a particular acoustic path can be combined with a reference ocean arrival time t_R to determine a *travel time anomaly*:

$$\Delta\tau_R = t_o - t_R . \tag{1}$$

The reference travel time is some predetermined standard derived by a computation of the travel time through the reference environment using a standard ray propagation model (e.g., Bellhop).

Any means of determining an observed travel time anomaly will include some measurement error, and we represent the errors associated with an observation as $\mathbf{e_z}(t)$:

$$\Delta\tau_{\mathrm{True}} = \Delta\tau_R + \mathbf{e_z}. \tag{2}$$

C. Model-predicted travel time anomalies

We assume that the path of an acoustic ray of interest only deviates slightly from the path resulting from the reference sound speed structure, c_R, at any time. Thus, temporal changes in travel time are primarily associated with changes in propagation speed along the reference medium path (the ''frozen ray'' approximation). If we know the path an acoustic ray would take through the model domain (individual grid cells denoted by $i=1,2,3,...,N$), we can calculate an estimated arrival time using

$$t_m = \Sigma[\Delta L_i/(c_{m,i}+U_{m,i})] \tag{3}$$

where the subscript *m* denotes model-predicted values, $c_{m,i}$ is the sound speed in the *i*th model grid cell, $\Delta L_{m,i}$ is the distance that the ray travels through the *i*th grid cell, and $U_{m,i}$ is the component of the three-dimensional current along a particular direction of interest responsible for effectively increasing or decreasing the sound speed. The model sound speed anomaly for each grid is

$$\Delta c_i = (c_{m,i}+U_{m,i}) - c_{R,i}, \tag{4}$$

where $c_{R,i}$ is based on reference *T* and *S* values.

Rearranging (4), substituting into (3), linearizing using $c_{Ri}^2 \gg \Delta c_i^2$, and rearranging the result in terms of a travel time anomaly (relative to the exact same t_R as in the expression for $\Delta\tau_R$) give

$$\Delta\tau_m(t) = -\sum_{i=1}^{N} \frac{\Delta L_i \Delta c_i(t)}{c_{R,i}^2} = \mathbf{b}\boldsymbol{\Delta}\mathbf{c}^T, \tag{5}$$

where *N* is the number of model grid cells through which the ray travels. On the very right-hand side of (5), we have represented the summation as the multiplication of two vectors, **b** being the vector of the constant $-\Delta L_i/c_{R,i}^2$ terms and $\boldsymbol{\Delta}\mathbf{c}$ being the vector of the time-varying terms $\Delta c_i(t)$.

As with the observed travel time anomaly, the model-predicted travel time has errors relative to the true value $\Delta\tau_{\mathrm{True}}$:

$$\Delta\tau_{\mathrm{True}} = \Delta\tau_{\mathrm{m}} + e_{\mathrm{DIS}}(t) + e_{\mathrm{LIN}}(t) + e_{\mathrm{TSU}}(t). \tag{6}$$

Errors due to the discretization process (representing $\Delta\tau_m$ by a summation over *N* grid cells as opposed to an integral over a continuum) are denoted by $e_{\mathrm{DIS}}(t)$. The errors associated with the linearization approximation ($c_{\mathrm{Ri}}^2 \gg \Delta c_{\mathrm{i}}^2$) are contained in $e_{\mathrm{LIN}}(t)$. The errors resulting from the differences between model-predicted sound speed (i.e., *T*, *S*, and *U*) and the true sound speed are denoted by $e_{\mathrm{TSU}}(t)$.

The reader should note that we assume that the measurement error, $e_z(t)$, is a random variable that is uncorrelated with the model error terms.

D. The tomographic expression

Equation (5) must be transformed to relate travel time anomalies to ocean model variables (*T*, *S*, and **U**). We express the sound speed in the ocean as a sum of a reference sound speed and a sound speed perturbation:

$$c = c_R + \frac{\partial c}{\partial T}\Delta T + \frac{\partial c}{\partial S}\Delta S + U = c_R + \delta c. \tag{7}$$

For local ocean temperatures and salinities and small temperature and salinity variations, we can approximate the two partial derivatives as

$$\frac{\partial c}{\partial T}\Delta T \approx 4.947\Delta T = \alpha\Delta T, \quad \frac{\partial c}{\partial S}\Delta S \approx 1.34\Delta S = \beta\Delta S,$$

where α has units of m/s/°C and β has units of m/s/ppt. We note that α and β are only gross approximations to $\partial c/\partial T$ and $\partial c/\partial S$, but their use is required to maintain the linearity of the set of expressions used in the assimilation process.

If the ocean model provides a reasonable first-guess of T, S, and U, we can rearrange (7) to give

$$\Delta c_i = \alpha \Delta T + \beta \Delta S + U \tag{8}$$

for every grid cell in the ocean model. We use (8) to transform (5):

$$\Delta \tau_m = \mathbf{b} \boldsymbol{\Delta} \mathbf{v}^{\mathbf{T}} \tag{9}$$

where now

$$\mathbf{b} = [-\Delta L_1/c_{R,1}^2 \;\; -\Delta L_1 \alpha/c_{R,1}^2 \;\; -\Delta L_1 \beta/c_{R,1}^2 \;\; \cdots \;\; -\Delta L_N/c_{R,N}^2 \;\; -\Delta L_N \alpha/c_{R,N}^2 \;\; -\Delta L_N \beta/c_{R,N}^2]$$

and the $\boldsymbol{\Delta}\mathbf{v}$ vector is

$$\boldsymbol{\Delta}\mathbf{v} = [U_1 \;\; \Delta T_1 \;\; \Delta S_1 \;\; \cdots \;\; U_N \;\; \Delta T_N \;\; \Delta S_N].$$

Each of the parameters in $\boldsymbol{\Delta}\mathbf{v}$ is the model-predicted variable relative to the reference value, all of which are known. Essentially, (9) is our tomographic relationship, relating acoustic information to T, S, and $\mathbf{U}$.

We let the total number of rays present in any set of P source-receiver transects be M. We define $\mathbf{z}(t)$ as the column vector of the M travel time anomaly measurements ($\Delta \tau_R$'s) at time t. We define the matrix $\mathbf{H}$ as that whose rows are the above $\mathbf{b}$. Since each ray path may not go through the same number of grid cells N, the number of columns in $\mathbf{H}$ will be three times the maximum of the N's ($N_{\max}$), and there can be a number of zero entries in $\mathbf{H}$. Finally, we define $\mathbf{x}$ as the $3 \times N_{\max}$ column vector with the $\boldsymbol{\Delta}\mathbf{v}$ for each grid cell through which a ray path travels.

We can relate all these matrices by the *measurement matrix residual equation*:

$$\mathbf{z}(t) - \mathbf{H}\mathbf{x}(t) = \text{residual}.$$

The goal of the data assimilation method is to minimize the residual. We note that, if the acoustic model grid is not coincident with the ocean model grid, the measurement matrix $\mathbf{H}$ must be premultiplied by an interpolation matrix that maps the acoustic model grid onto the ocean model grid. In this study, the two grids coincide.

E. Assimilation of tomographic information into the ocean model

In our assimilation formulation, we define ocean parameters throughout the ocean model grid space as $\mathbf{x}$, $\mathbf{x}^{\mathbf{F}}$, and $\mathbf{x}^{\mathbf{A}}$ as vectors representing the *true state*, the *forecasted estimate* prior to assimilation, and the *analysis estimate* (after assimilation), respectively. The vectors $\mathbf{x}$, $\mathbf{x}^{\mathbf{F}}$, and $\mathbf{x}^{\mathbf{A}}$ are time dependent, and the three-dimensional T, S, and U fields form our state vector.

The basic expression we will use to determine the analysis (updated) field combines the forecast estimate $\mathbf{x}^{\mathbf{F}}$ with the acoustic-related measurements $\mathbf{z}_R$ (the $\Delta \tau_R$'s) using the model-related measurement matrix $\mathbf{H}$ as follows (the PSAS formulation):[5]

$$\mathbf{x}^{\mathbf{A}} = \mathbf{x}^{\mathbf{F}} + \mathbf{K}(\mathbf{z}_R - \mathbf{H}\mathbf{x}^{\mathbf{F}}). \tag{10}$$

As mentioned above, $\mathbf{z}_R - \mathbf{H}\mathbf{x}^{\mathbf{F}}$ is the measurement residual, the difference between the observation and the forecasted variable. $\mathbf{K}$ is the residual (Kalman) gain matrix:

$$\mathbf{K} = \mathbf{P}_{\mathbf{F}}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{P}_{\mathbf{F}}\mathbf{H}^{\mathbf{T}} + \mathbf{R})^{-1}.$$

$\mathbf{P}_{\mathbf{F}}$ is the spatial covariance of errors in the forecast, and $\mathbf{R}$ is the spatial covariance of errors in the observations. The formulations for these error matrices will be presented in the following section. But it is readily seen that, if $\mathbf{R} \gg \mathbf{H}\mathbf{P}_{\mathbf{F}}\mathbf{H}^{\mathbf{T}}$, the Kalman gain $\mathbf{K}$ approaches zero: $\mathbf{x}^{\mathbf{A}} = \mathbf{x}^{\mathbf{F}}$. Also, the Kalman gain distributes the measurement residual throughout the forecast model domain. The matrix $\mathbf{P}_{\mathbf{F}}$ can have values for each grid cell, and, thus, $\mathbf{K}$ will have a value for each grid cell, even for just one source-receiver pair. This is a significant advantage over inversion tomography where the covariance matrix tends to be limited to just the region being acoustically illuminated. Typically, corrections based on observations are assigned mostly to regions closest to the observations and areas where the forecast model error is the highest.

Note that $\mathbf{H}$ is used when determining $\mathbf{K}$ and during the assimilation process expressed in (10). Calculating $\mathbf{K}$ is performed prior to the assimilation process. Thus, even though we could use model-predicted T and S to obtain better estimates for $\partial c/\partial T$ and $\partial c/\partial S$ when employing (10), we do not have these values when calculating $\mathbf{K}$. This is the reason we employ the gross approximations of $\partial c/\partial T = \alpha$ and $\partial c/\partial S = \beta$, to make $\mathbf{H}$ consistent between the process of calculating $\mathbf{K}$ and then later using (10).

IV. SPECIFYING THE ERROR COVARIANCE MATRICES

For this preliminary study, (10) was simplified by assuming that travel time anomalies were primarily a result of the differences between predicted and ocean water temperatures. In this case, $\mathbf{x}^{\mathbf{F}}$ and $\mathbf{x}^{\mathbf{A}}$ are column vectors of forecasted and analysis temperatures relative to the monthly reference temperatures: $\mathbf{x}^{\mathbf{F}} = \mathbf{T}_{\text{model}} - \mathbf{T}_{\text{reference}}$ and $\mathbf{x}^{\mathbf{A}} = \mathbf{T}_{\text{analysis}} - \mathbf{T}_{\text{reference}}$. Since $\mathbf{T}_{\text{reference}}$ appears on both sides of (10), we see that the expression reduces to

$$\mathbf{T}_{\text{analysis}} = \mathbf{T}_{\text{model}} + \mathbf{K}(\mathbf{z}_R - \mathbf{H}\mathbf{x}^{\mathbf{F}}). \tag{11}$$

The rows of the $\mathbf{H}$ matrix now consist of

$$\mathbf{b} = [-\Delta L_1 \alpha/c_{R,1}^2 \;\; -\Delta L_2 \alpha/c_{R,1}^2 \;\; \cdots \;\; -\Delta L_N \alpha/c_{R,N}^2]$$

for each ray path. As before, $\mathbf{z}_R$ is the column vector of ''observed'' travel time anomalies. Thus, if $\mathbf{K}$ is properly determined, all the terms on the rhs of (11) are defined, and we can solve for $\mathbf{T}_{\text{analysis}}$.

A. Calculating the covariance functions

The modification of the model variables results from the application of (10). As such, the magnitude of the change of a model variable is critically dependent on the error covariance matrices. The model error covariance matrix is defined as

$$\mathbf{P}_{\mathbf{F}} = E[(\mathbf{x} - \mathbf{x}^{\mathbf{F}})(\mathbf{x} - \mathbf{x}^{\mathbf{F}})^T].$$

Of the various elements of $\mathbf{P_F}$, the e_{TSU} errors (differences between the model and the true ocean T, S, and $\mathbf{U}$) will likely be the most dominant. Typical values of sound speed in water and the magnitude of the rms errors determined from model/data comparisons (Sec. II B) indicate that the linearization error, e_{LIN}, is negligible: the value of c_{Ri}^2 will be at least four to five orders of magnitude larger than Δc_{i}^2, even with substantial errors in salinity. Numerical simulations suggest that discretizing the integral on the gridding scales used by the hydrodynamic model results in $e_{\mathrm{DIS}} \leqslant 0.1$ ms for rays that interact with the surface one to three times. Thus, we ignore e_{DIS} and concentrate on developing a reasonable approximation for e_{TSU}.

The most accurate method of calculating the model error covariance matrix $\mathbf{P_F}$ is to use many realizations in space and time of model-predicted variables along with corresponding observations. Typically, such observations are nonexistent. One approach to deal with the lack of model-data comparisons is to estimate $\mathbf{P_F}$ by the spatial covariance from the time sequence of the model-predicted variable over a set of (x,y,z) locations:[16]

$$\mathbf{P_F} = E[(\mathbf{T}_{\mathrm{MODEL}} - E[\mathbf{T}_{\mathrm{MODEL}}])(\mathbf{T}_{\mathrm{MODEL}} - E[\mathbf{T}_{\mathrm{MODEL}}])^T]. \quad (12)$$

It is easily shown that the model covariance equals the sum of the model error covariance plus the covariance of the true state of the ocean. Thus, $\mathbf{P_F}$ given by (12) always results in overestimating the error covariance. As a result, the rate of spatial decorrelation of the errors of a variable can be underestimated. Therefore we would expect the observations to be more spatially limited in their impact on the analysis fields, which in itself is not an adverse consequence but may not make full use of the observations.

A second method involves estimating the error covariance matrix $\mathbf{P_F}$ utilizing scaled, time histories of model-predicted temperatures. In the first step of this method, a time series of errors is specified for a given temperature $T(x,y,z,t)$ as

$$T_{\mathrm{error}}(x,y,z,t) = T(x,y,z,t+24 \text{ hours}) - T(x,y,z,t).$$

Using the 24-h offset, the ''errors'' are simply a result of day-to-day variability in atmospheric forcing, the phase of the tides, wave conditions, and the MODAS T-S fields. From this we can calculate n days worth of estimates of errors (for our study, errors at hourly intervals) for the ocean variable T. The second step is the scaling of the errors T_{error}. This scaling process relies on the existence of model predictions and corresponding observations with which to calculate an rms model error. In our case, 175 bathythermographs were used to calculate an rms error for T as a function of depth. Thus, each $T_{\mathrm{error}}(x,y,t)$ for a specific z can be scaled [each value of $T_{\mathrm{error}}(x,y,z,t)$ increased or decreased] so that the model-model rms errors match the observed-model rms errors. This scaling allows us to produce estimates between the model and the true state of the ocean that are realistic in that the rms model-model differences match rms model-data differences. We use the scaled time histories to calculate the appropriate spatial covariance matrix using

$$\mathbf{P_F} = E[(\mathbf{T}_{\mathrm{error}} - E[\mathbf{T}_{\mathrm{error}}])(\mathbf{T}_{\mathrm{error}} - E[\mathbf{T}_{\mathrm{error}}])^T]. \quad (13)$$

Another consideration is the length of the time series over which the spatial covariance is calculated. Commonly, the spatial correlation structure varies with factors such as the phase of the internal tides. Thus, the error covariance matrices should be recalculated on a fairly regular basis. The time span over which $\mathbf{T}_{\mathrm{MODEL}}$ or $\mathbf{T}_{\mathrm{error}}$ extend should partially reflect conditions during which travel time anomalies are to be assimilated. If the fortnightly phase of the internal tide is the primary factor controlling spatial structure, then a time series of ~3 days may be appropriate. In this study, we consider estimates of the model covariance matrix using 3 days of hourly data.

The observation error covariance matrix is defined as

$$\mathbf{R} = E[(\mathbf{e_z} + E[\mathbf{e_z}])(\mathbf{e_z} + E[\mathbf{e_z}])^T].$$

There are two components of the measurement error vector $\mathbf{e_z}$. The first component is given by

$$\sigma_t = \frac{1}{bw\sqrt{\mathrm{SNR}}},$$

where bw is the bandwidth of the signal ($bw = 2\pi\Delta f$) and SNR is the signal-to-noise ratio. In our experiment performed at PMRF during June–July 2003, we had a bandwidth of 3000 Hz. Assuming a signal-to-noise ratio of 10 dB, this error is around 32 μs.

The second component of $\mathbf{e_z}$ is due to the fluctuations in the water column temperature and salinity, the roughness of the ocean surface, and uncertainties in the bathymetry. One approach to estimate this error is from the acoustic data itself. This is done by estimating the standard deviation of the arrivals for a set of pings between a source-receiver pair. This information is presented in the following section.

B. Implementation

In implementing our data assimilation scheme, we limited the model grid cells impacted by acoustic observations to those within 20 km of any model grid cell through which ray paths being considered traveled. This reduced the number of covariance functions that had to be calculated. If $L_{\max}$ is the number of grid cells within the 20-km range, then $\mathbf{P_F H}^T$ for the p ray is an $L_{\max}$ column vector. There are 12 monthly column vectors ($\mathbf{c_R}$ for a given grid cell varies by month) for the any ray path. All the elements were calculated and stored in a database. Similarly, all terms in the $\mathbf{HP_F H}^T$ matrix are known, and each monthly matrix was calculated. The monthly $\mathbf{HP_F H}^T$ matrices were inverted and multiplied by $\mathbf{P_F H}^T$ to give 12 $L_{\max} \times P_{\max}$ arrays, where $P_{\max}$ is the number of ray paths being considered (i.e., there are $P_{\max}$ arrival time anomalies).

Thus, the analysis and assimilation software only requires the monthly databases of (1) the $\mathbf{P_F H}^T(\mathbf{HP_F H}^T)^{-1}$ elements, (2) the reference temperatures along the acoustic paths being considered [for Δv in (9)], and (3) the $-\alpha \Delta \mathrm{L_N}/c_{R,N}^2$ values [for $\mathbf{b}$ in (9)].

As in any assimilation scheme, there are additional factors that have been incorporated in our process. First, we

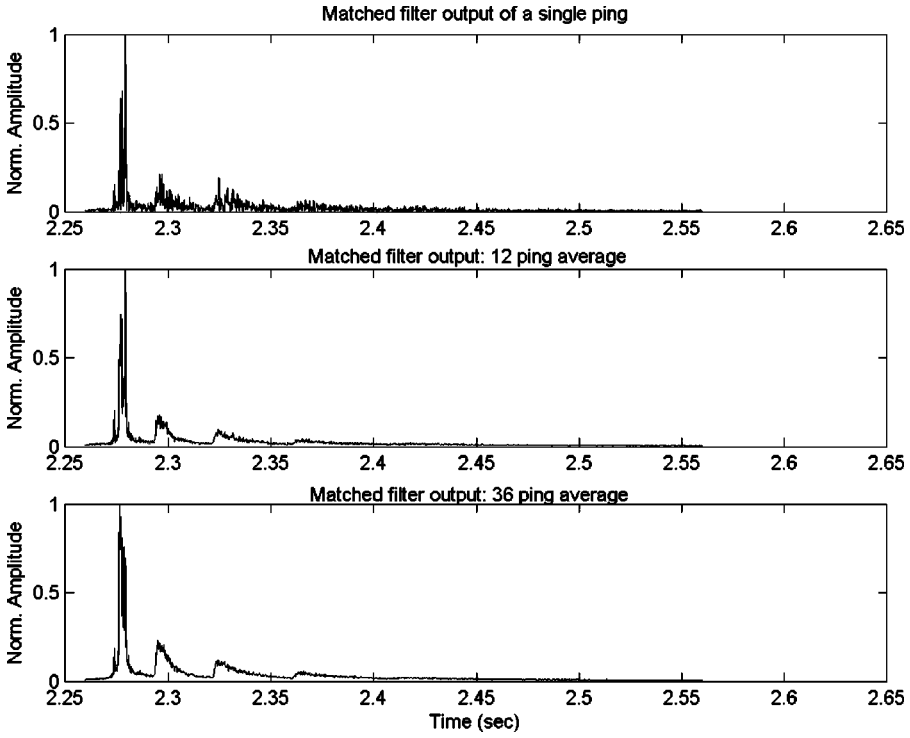


FIG. 2. Matched filter output for a single transmission (top), an average of 12 transmissions (middle), and an average of 36 transmissions (bottom).

assumed that the MODAS temperature field, $T_M(x,y,z)$, reflects a norm for the $\mathbf{T}_{\text{analysis}}(x,y,z)$ field. We used the 175 bathythermographs (Sec. II B) to calculate relative ranges for each vertical level of the ocean model: $Ra(z)=T(z)_{\max}-T(z)_{\min}$. We then limited the replacement of the elements in the $\mathbf{T}_{\text{model}}$ field to those grid cells in which the assimilation process resulted in (a) a water temperature within the range of $T_M(x,y,z)\pm Ra(z)/2$, (b) a water temperature closer to $T_M(x,y,z)-Ra(z)/2$ when $\mathbf{T}_{\text{model}}<T_M(x,y,z)-Ra(z)/2$, or (c) a water temperature closer to $T_M(x,y,z)+Ra(z)/2$ when $\mathbf{T}_{\text{model}}>T_M(x,y,z)+Ra(z)/2$. These limitations were found to aid in preventing the assimilation process from making model-predicted temperatures that were already too cold (warm) from being made even colder (warmer). This is not to say that other forcings cannot result in model-predicted temperatures that drift away from the $T_M(x,y,z)\pm Ra(z)/2$ field. Only the assimilation process is constrained to alter the model-predicted temperatures toward a range about the MODAS temperatures.

V. THE ACOUSTIC DATA

Each signal transmitted from a PMRF source is acquired by an acoustic data acquisition system (ADAS) at designated receivers. The transmitted signal is then replica-correlated with the received signal. The received signal is modeled as a sum of ray arrivals given by

$$r(t)=\sum_n a_n s(t-\tau_n),$$

where a_n is the weight associated with each arrival, $s(t)$ is the transmitted signal, and τ_n is the delay associated with each arrival. When correlated with the transmitted signal, the output will have peaks at times corresponding to the travel times of eigenrays between the source and receiver. There are two sources of error in the estimation of travel time if we use the above representation. First, the sound channel is dispersive, and we cannot expect the transmitted signal shape to remain unchanged as it propagates through the channel. The second error is because we have not correctly modeled phase changes that occur during propagation (due to caustics or boundary reflections). Simulations show that the errors in arrival time estimates due to both these causes are of the order of tens of microseconds, which is less than the error in arrival time estimates as a result of the finite source bandwidth. Thus, these errors can be ignored for our application.

Two candidate transmit signals were tested: a linear chirp and a 511-digit *m*-sequence. In areas where the acoustic signal interacts with waves on the ocean surface, there can be a Doppler shift that compresses or elongates the signal envelope, depending on the direction of motion of the ocean surface relative to the incident acoustic energy. Matched filter output of this Doppler-shifted signal gives rise to errors in (a) estimating the arrival time and (b) the amplitude of the matched filter output of each arrival. Analysis of the errors in estimating arrival times showed that the errors in the phase-coded sequence were less than that of the chirp signal. However, the amplitude of the matched filter output remained practically unaltered in the case of chirp signal, while in the phase-coded sequence it is reduced substantially. This resulted in greater difficulties in detecting a phase-coded signal in the presence of background noise. Since averaging can eliminate the error in the arrival time due to surface motion, the transmissions used in this study were chirp signals with a center frequency of 9.5 kHz and a bandwidth of 3 kHz.

To improve the signal-to-noise ratio and to reduce the impact of surface motion, we averaged over a number of acoustic transmissions. Under normal circumstances, it would have been appropriate to send a large train of acoustic

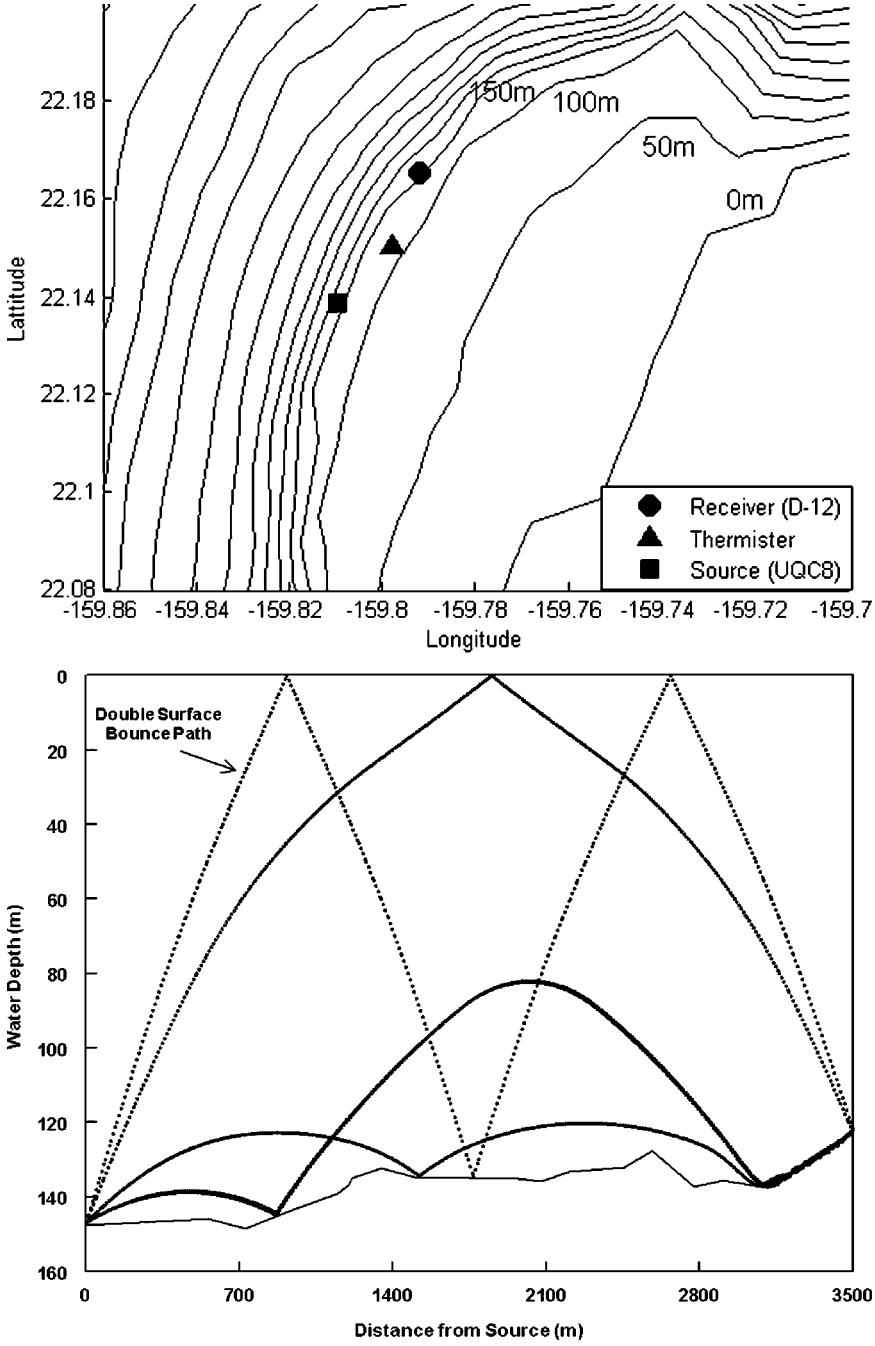


FIG. 3. Top panel—bathymetry and locations of the source (*), receiver (○), and where thermistor data were collected (△). Bottom panel—examples of ray paths between the source and receiver.

pulses and perform an average over this train of pulses. However, this was not possible because of ''cross talk'' between the transmitter and receivers at PMRF. Instead, we transmitted a sequence of 12 acoustic transmissions that consisted of four groups separated by 8 s. Each acoustic transmission had a duration of 0.1 s, with an interval of 0.4 s between each transmission. The number of pulses in a group was restricted to three transmissions to avoid interference due to the cross talk. The distance to the nearest receiver set this limitation. The maximum distance between the source and selected receivers dictated the 8-s separation between groups of transmissions. The string of 12 transmissions was repeated three times with an interval of about 30 s. The length of each train of pulses (i.e., about 28 s) and the time interval between each train were selected on the basis of the requirements of the data acquisition system.

The matched filter outputs for a varying number of transmissions are shown in Fig. 2. The top panel shows the output using only one transmission. The output using the average of 12 transmissions is shown in the middle panel, and the output using the average of 36 transmissions is shown in the lower panel. We see that a considerable enhancement of the signal-to-noise ratio is achieved by averaging over 36 transmissions.

The arrival structure in Fig. 2 consists of one stronger arrival followed by three weaker arrivals. An eigenray[17] analysis for this particular source/receiver pair was performed using a sound speed field for the region obtained from the ocean model. Based on this analysis, the earliest group of arrivals consists of rays that travel from source to receiver without interacting with either the ocean surface or bottom and others that include a ray that interacts with the

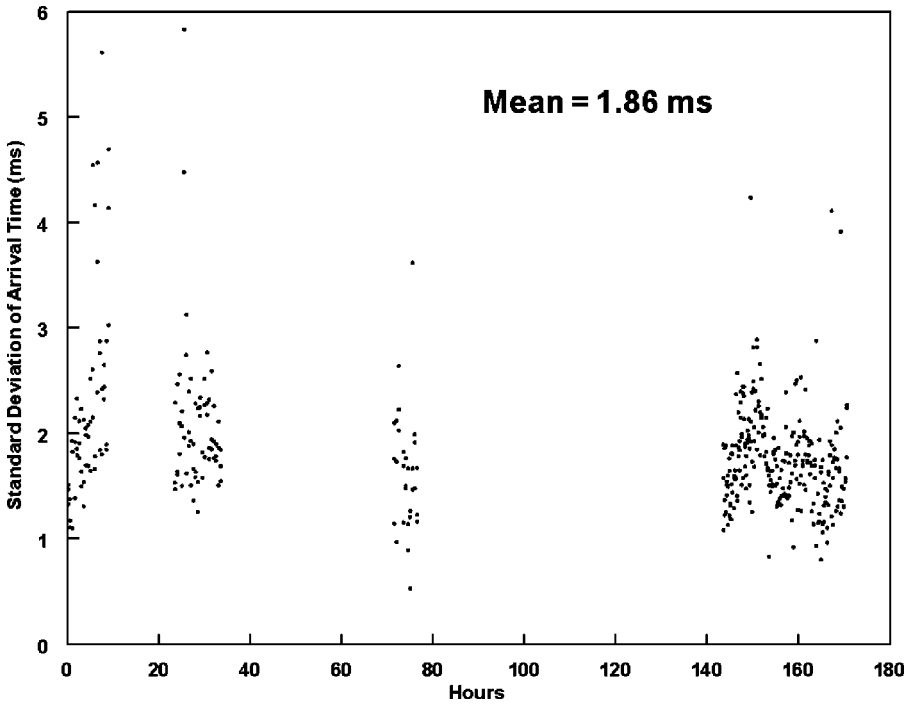


FIG. 4. Standard deviation of the arrival times for acoustic data between PMRF hydrophones D9 and D12 during late June and early July 2003 at PMRF.

surface only once and rays that hug the bottom and have repeated interactions with the bottom. The arrival times of rays that have only interactions with the bottom carry little information about the bulk of the water column. This plus the problem of delineating individual ray arrival times within the first group of arrivals lead us to neglect these arrivals in our tomography analysis.

The subsequent three arrivals in Fig. 2 correspond to rays that have two, three, and four surface bounces, respectively. As mentioned before, we concentrated on analyzing the path that bounced off the ocean surface only twice due to the drop in signal-to-noise for higher multiple surface bounce acoustic paths (the third and fourth arrivals).

For the initial development and testing of this technology, we worked with just one ray path. The ADAS was used to collect travel time observations for a double-surface bounce ray path at PMRF just offshore of the 90-m isobath (Fig. 3). The ray path was between a bottom-mounted source (D9) some 3.5 km from a bottom-mounted receiver (D12). From an eigenray analysis performed using a mean sound speed structure for the area, the arrival time of a ray with two surface bounces is about 2.3 s. This is used to readily identify the arrival time of the two-surface bounce ray in the ADAS data (Fig. 2). An enhanced view (not shown) of the

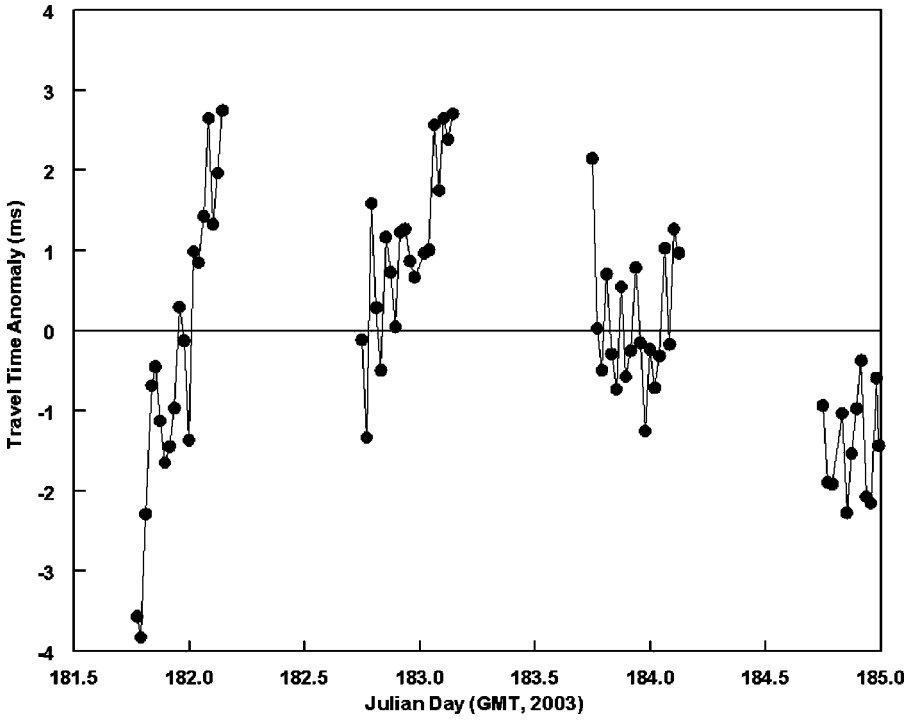


FIG. 5. Calculated travel time anomalies for a two-surface bounce ray path over a 3.5-km distance between the PMRF source D9 and receiver D12.

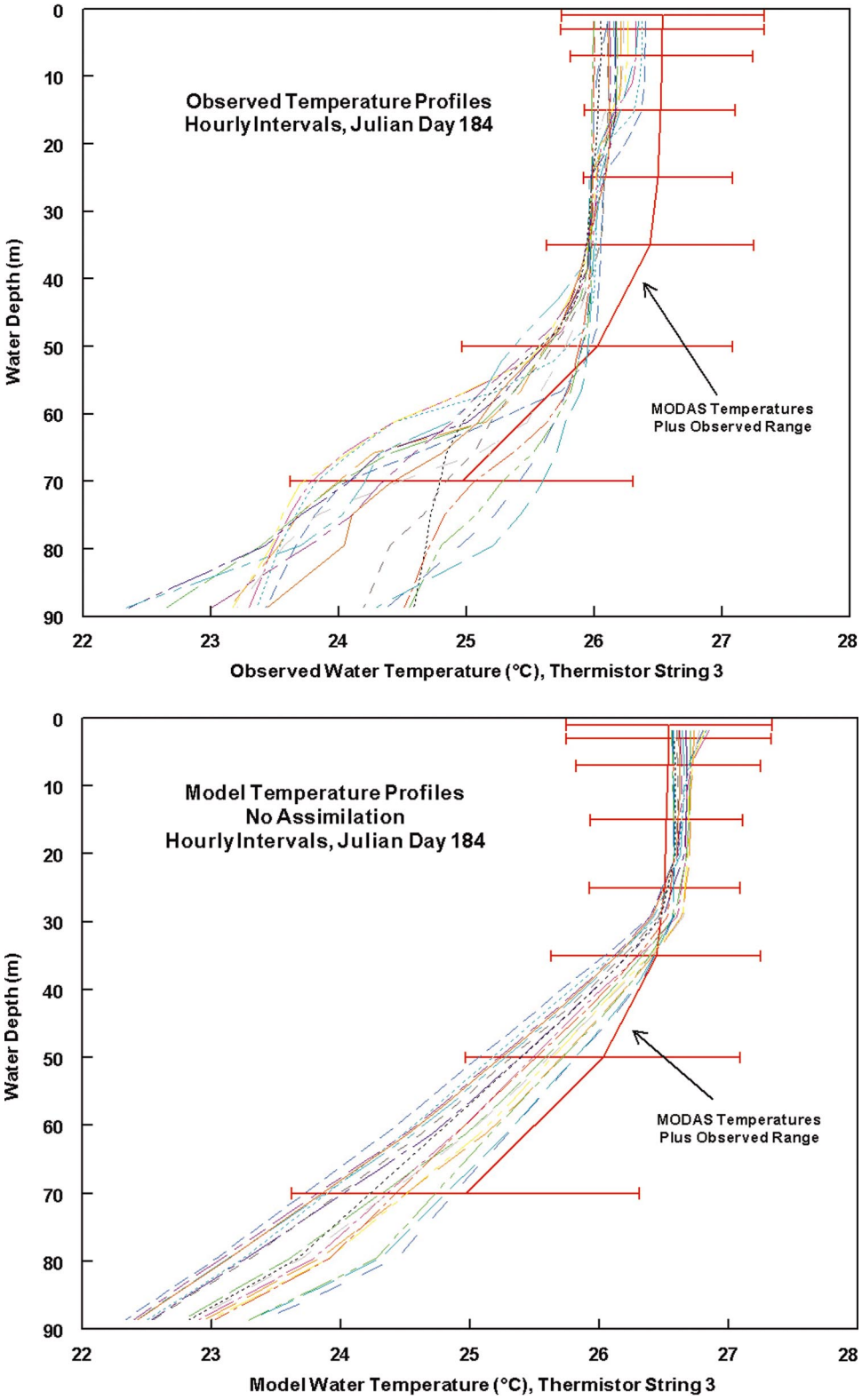


FIG. 6. Hourly temperature profiles for 3 July 2003 at the thermistor site shown in Fig. 3: observed (top) and predicted, no assimilation (bottom). The MODAS temperature profiles for that date and location are also shown (solid red), along with the observed ranges of temperatures (horizontal solid red) determined from the 178 BTs, centered on the MODAS temperatures.

matched filter output corresponding to the arrival time for this ray indicates multiple peaks that are likely the result of the roughness of the ocean surface and the ocean bottom. In order to determine the arrival time of a particular ray, we performed cluster analysis of the arrival times and their relative amplitudes of all arrivals that represented the acoustic ray of interest. The centroid of the cluster of peaks was used as the estimate of the arrival time. In addition, the variance of the cluster of peaks (arrival times) was used to estimate the error covariance matrix **R**.

The data were used to estimate the $\mathbf{e}_z$ error term for the two-surface bounce ray between D9 and D12. The standard deviations of the all arrivals for sets of pings covering a 180-h period are shown in Fig. 4. As can be seen, there is a significant amount of variability in the standard deviation, varying from 0.5 to 6 ms. The mean was 1.8 ms. We note

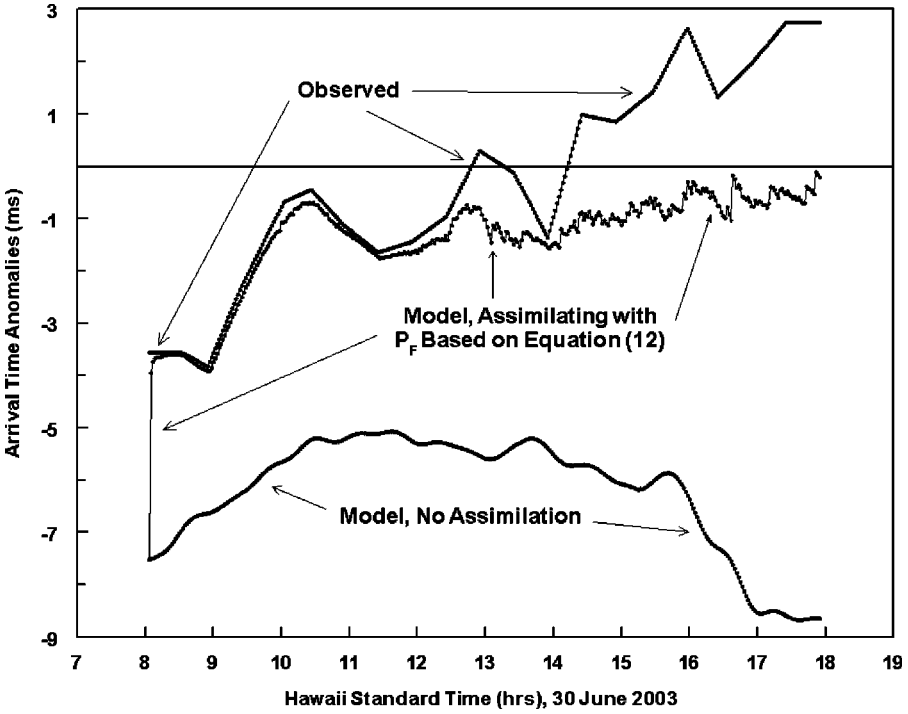


FIG. 7. Observed travel time anomalies ($\mathbf{z}_R$) and model-predicted travel time anomalies ($\mathbf{Hx}^F$) using (12). Each dot represents one time step (100 s) of the numerical ocean model. The variations of $\mathbf{Hx}^F$ with $\mathbf{K}=0$ (no assimilation) are shown by the bottom curve.

that this $\mathbf{e}_z$ error is orders of magnitude greater than the σ_t error due to noise in the signal and the limited bandwidth of the signal (32 μs). Thus, we can neglect σ_t.

VI. DATA ASSIMILATION TEST CASE, JULY 2003

During an Office of Naval Research experiment[18] at PMRF during June–July 2003, a series of thermistor strings were placed along the 90-m isobath just shoreward of the D9-D12 source-receiver pair. The distance between the D9-D12 ray path and the thermistor arrays was ~0.5 km. As such, the thermistor data provide a means of assessing the impact of assimilating the D9-D12 arrival time anomaly data into the ocean circulation model.

During 30 June–3 July 2003, the ADAS was used to generate 8–11-kHz chirps (linear FM sweep) with a duration of 100 ms from D9. The signal was transmitted 36 times over a 2.5-min period every half hour. Data were collected from about 8:00 AM to about 5:00 PM local time (1800 GMT to 0300 GMT) for all 4 days. Output of each transmission in the three chirp trains was run through the matched filter process, and then the 36-transmission average was calculated. In

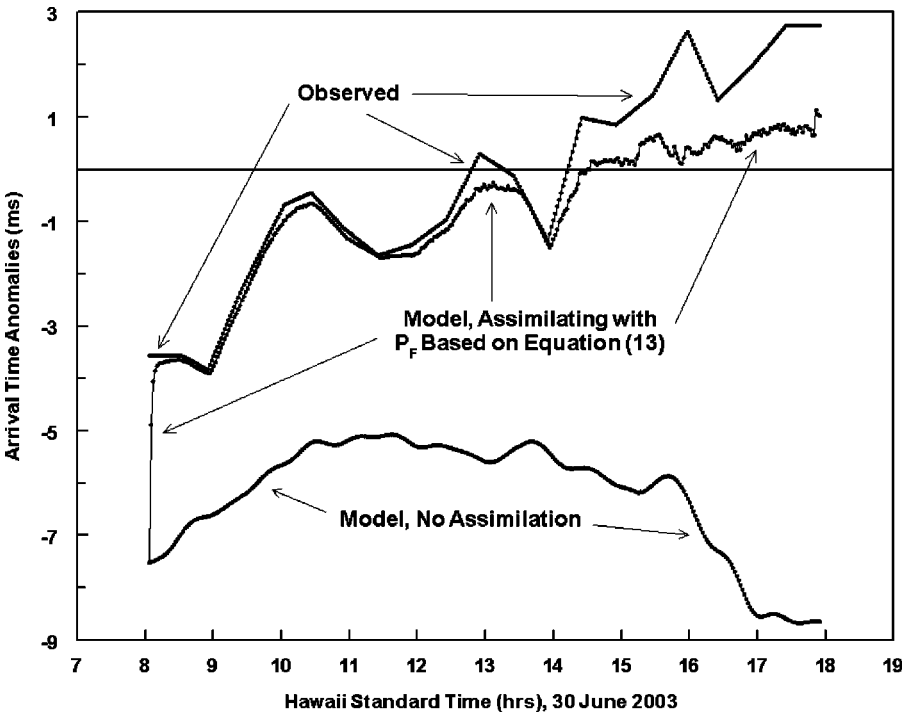


FIG. 8. Observed travel time anomalies ($\mathbf{z}_R$) and model-predicted travel time anomalies ($\mathbf{Hx}^F$) using (13). Each dot represents one time step (100 s) of the numerical ocean model. The variations of $\mathbf{Hx}^F$ with $\mathbf{K}=0$ (no assimilation) are shown by the bottom curve.

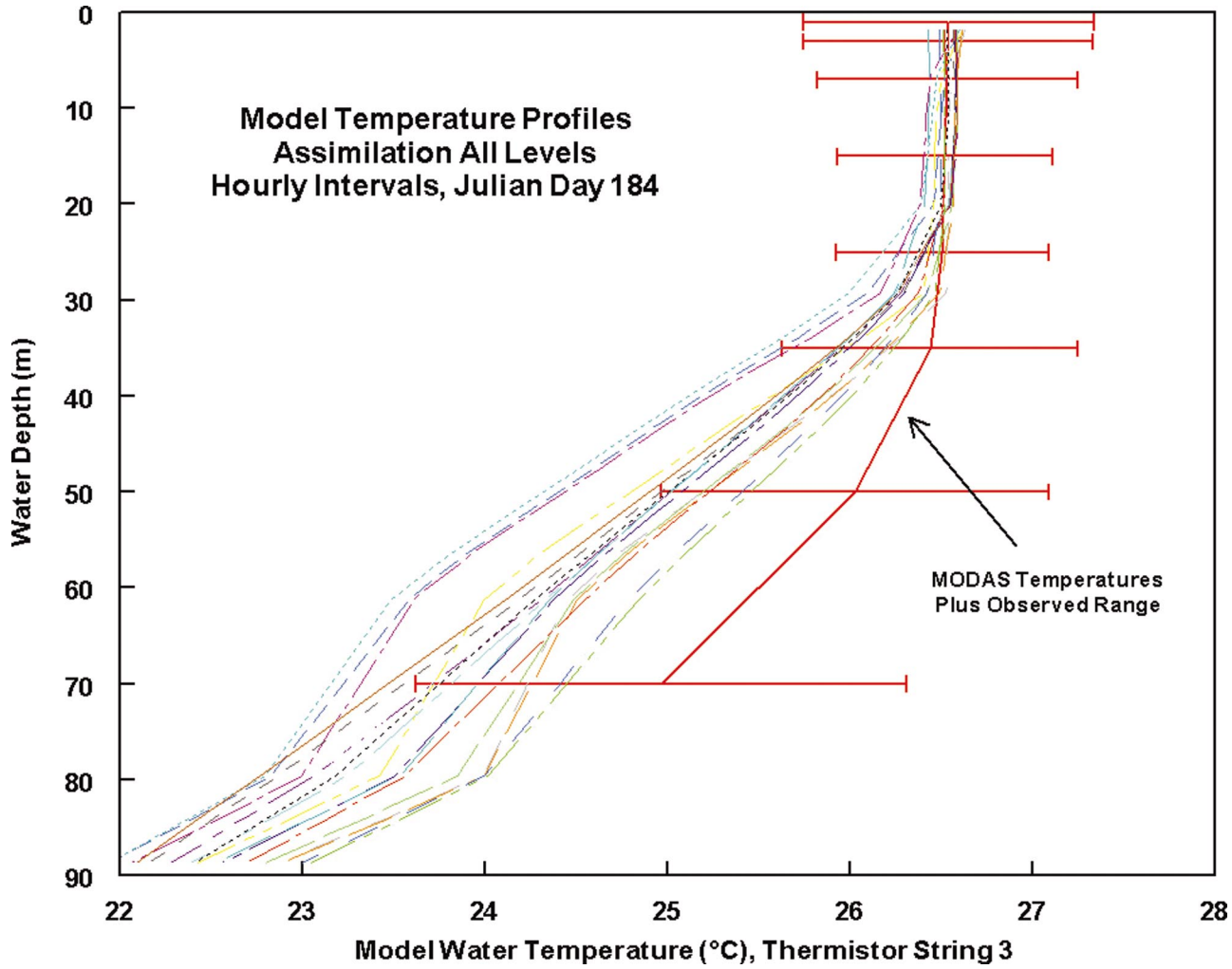


FIG. 9. Model-predicted hourly temperature profiles for 3 July 2003 at the thermistor site shown in Fig. 3. The MODAS temperature profiles for that date and location are also shown (solid red), along with the observed ranges of temperatures (horizontal solid red) determined from the 178 BTs, centered on the MODAS temperatures.

cases where multiple peaks resulted in an ambiguity in determining the arrival time, the averaged matched filter output was low-pass filtered.

After determining all the arrival times of the double surface bounce path between D9 and D12, travel time anomalies were calculated for assimilation into the ocean model. These are shown in Fig. 5. There are distinct longer-term variations in arrival time anomalies that are likely a result of tidal fluctuations and/or surface heating and cooling. But the observed travel time anomalies also have large hour-to-hour variations. These short-term fluctuations could be the result of signal processing (i.e., the ''noise'' seen in the curves in Fig. 2) or actual ocean processes that alter the sound speed of the water column. When translated to geophysical fluid dynamics, the short-term variations mean short space scale phenomena: e.g., solitons. Even if these fluctuations are not a result of signal processing, our ocean model does not have the horizontal resolution (tens of meters) that would be required to reproduce such short space scale phenomena. To mitigate the impact of such fluctuations in the arrival time anomalies (which would generate spurious gravity waves in the ocean model), we limit the magnitude of any changes to the model-predicted temperatures to be $<2.7\times10^{-4}$ °C/s for each time step. This would still allow up to a 5–6 °C temperature change over a 6-h period, a not-uncommon signature of internal tides around the Hawaiian Islands.

The ocean model was first executed without any acoustic data assimilation, and the model-predicted water temperatures were compared to the thermistor data. A comparison of the model-predicted and observed water temperatures at the site shown in Fig. 3 on 3 July 2003 is shown in Fig. 6. There is a distinct bias in the surface mixed layer (observed temperatures >26 °C) where the model-predicted temperatures are too warm. For the cooler waters at depth, the scatter is considerable, with the model predictions being as much as about 0.8 °C too warm and 2.2 °C too cool.

A. Tests of assimilation parameters

We used $\mathbf{e_z}=1.86$ ms, or $\mathbf{R}=(1.86\times10^{-3}\,\mathrm{s})^2=3.46\times10^{-6}\,\mathrm{s}^2$. For estimating $\mathbf{P_F}$, we have put forth two possible methods: those in Eqs. (12) and (13). As a measure of quantifying the magnitude of the impact of assimilating the acoustic information for different $\mathbf{P_F}$'s, we compared the two quantities of the measurement residual, $\mathbf{z}_R$ and $\mathbf{Hx}^\mathbf{F}$. Recall that $\mathbf{z}_R$ is the observed travel time anomaly (Fig. 5) and $\mathbf{Hx}^\mathbf{F}$

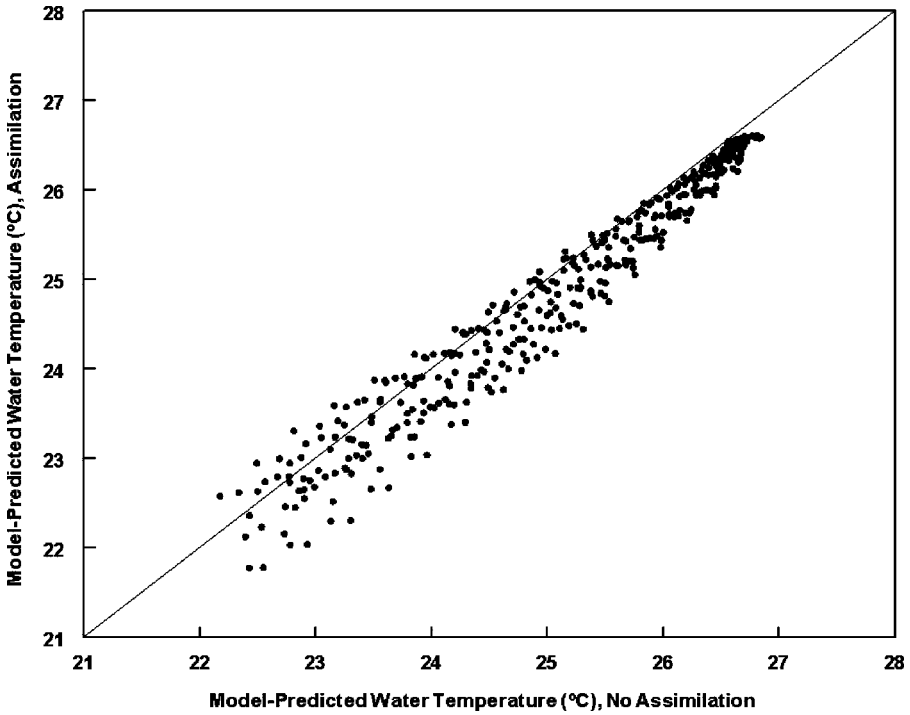


FIG. 10. A scatter plot of model-predicted water temperatures at one of the thermistor strings for 3 July 2003, with (ordinate) and without (abscissa) assimilation. Points below the diagonal line indicate that the assimilation process resulted in cooling.

is the corresponding travel time anomaly predicted by the ocean model. If the effect of the assimilation is significant, then $\mathbf{Hx^F} \rightarrow \mathbf{z}_R$ over time.

We first used $\mathbf{P_F}$ as defined in Eq. (12) with the time series used to calculate $\mathbf{P_F}$ being 72 h of temperatures for each model grid cell. An example of the variations of $\mathbf{Hx^F}$ and $\mathbf{z}_R$ is shown in Fig. 7. Each dot in Fig. 7 represents one time step of the ocean model (100 s). When the assimilation process started, $\mathbf{Hx^F}$ was approximately twice as large as $\mathbf{z}_R$, with about a 4-ms difference between the two. Within three time steps (and, thus, three new sets of $\mathbf{T}_{\text{analysis}}$ ingested by the model), $\mathbf{Hx^F}$ was reproducing $\mathbf{z}_R$ quite well. This good reproduction continued for the next 3–4 h. By about 12:45 hours local time, the difference between $\mathbf{z}_R$ and $\mathbf{Hx^F}$ began to increase, reaching a maximum of the order of 3+ ms by 18:00 hours local time.

A similar test was conducted using $\mathbf{P_F}$ as defined in (13), and the results are shown in Fig. 8. As before, the assimilation initially resulted in a good reproduction of $\mathbf{z}_R$. But, in this case, the good reproduction lasted longer, for about 6.5 h. By 15:00 hours local time, the differences between $\mathbf{z}_R$ and $\mathbf{Hx^F}$ became larger but were noticeably smaller than those shown in Fig. 7. By 18:00 hours, the differences were of the order of 2 ms.

We note that, as the differences between $\mathbf{z}_R$ and $\mathbf{Hx^F}$ in Figs. 7 and 8 increase, $\mathbf{Hx^F}$ is characterized by higher frequency oscillations. This type of variability commonly indicates that other model forcing (e.g., tidal, atmospheric, etc.) is substantial enough to counter the adjustments resulting from assimilating the observed travel time anomalies, thus explaining the increasing differences between $\mathbf{z}_R$ and $\mathbf{Hx^F}$. The bottom curves in Figs. 7 and 8 show $\mathbf{Hx^F}$ with no assimilation. It is obvious that one of the other forcings is pushing $\mathbf{Hx^F}$ to become more negative after about 14:00 hours, with a strong negative trend after 15:30 hours. This is opposite to the trend of $\mathbf{z}_R$.

Based on the results shown in Figs. 7 and 8, we use Eq. (13) for defining $\mathbf{P_F}$ for our assimilation exercises. In doing so, we calculated the value of $\mathbf{HP_FH}^T$ for the single acoustic path between D9 and D12. This value was $7.01\times10^{-6}\,\text{s}^2$. Thus, $\mathbf{HP_FH}^T$ was about twice that of $\mathbf{R}\,(3.46\times10^{-6}\,\text{s}^2)$.

B. Impact of assimilation process, 3 July 2003

Simulations were performed with the assimilation procedure executed whenever there were any arrival time anomaly data within 0.5 h of the model simulation time. The $\mathbf{P_F}$ matrix was determined using (13) with 3 days of hourly model results. In Fig. 9 we show the model temperatures at the same thermistor string whose data are shown in Fig. 6 (top panel). We see that the assimilation process resulted in cooler surface waters, but the warmer surface bias still exists. It is apparent that the assimilation lowered the subsurface water temperatures also (observed temperatures$<26\,^\circ$C), with the maximum overprediction in deeper waters being reduced to $+0.5\,^\circ$C. However, the model temperatures in the lower part of the water column were lowered too much, with underpredictions being now as large as $-2.3\,^\circ$C.

The overall impact of the data assimilation is better depicted in Fig. 10. In general, the temperatures of the water column were lowered (points below the diagonal line), with the cooling being larger in magnitude as the depth increased (toward the left-hand side of Fig. 10). We also see that the assimilation process resulted in some increasing temperatures (points above the diagonal) further down in the water column. The assimilation process induced relatively modest temperature changes for the 24-h period of 3 July 2003, ranging from $+0.5$ to $-0.8\,^\circ$C.

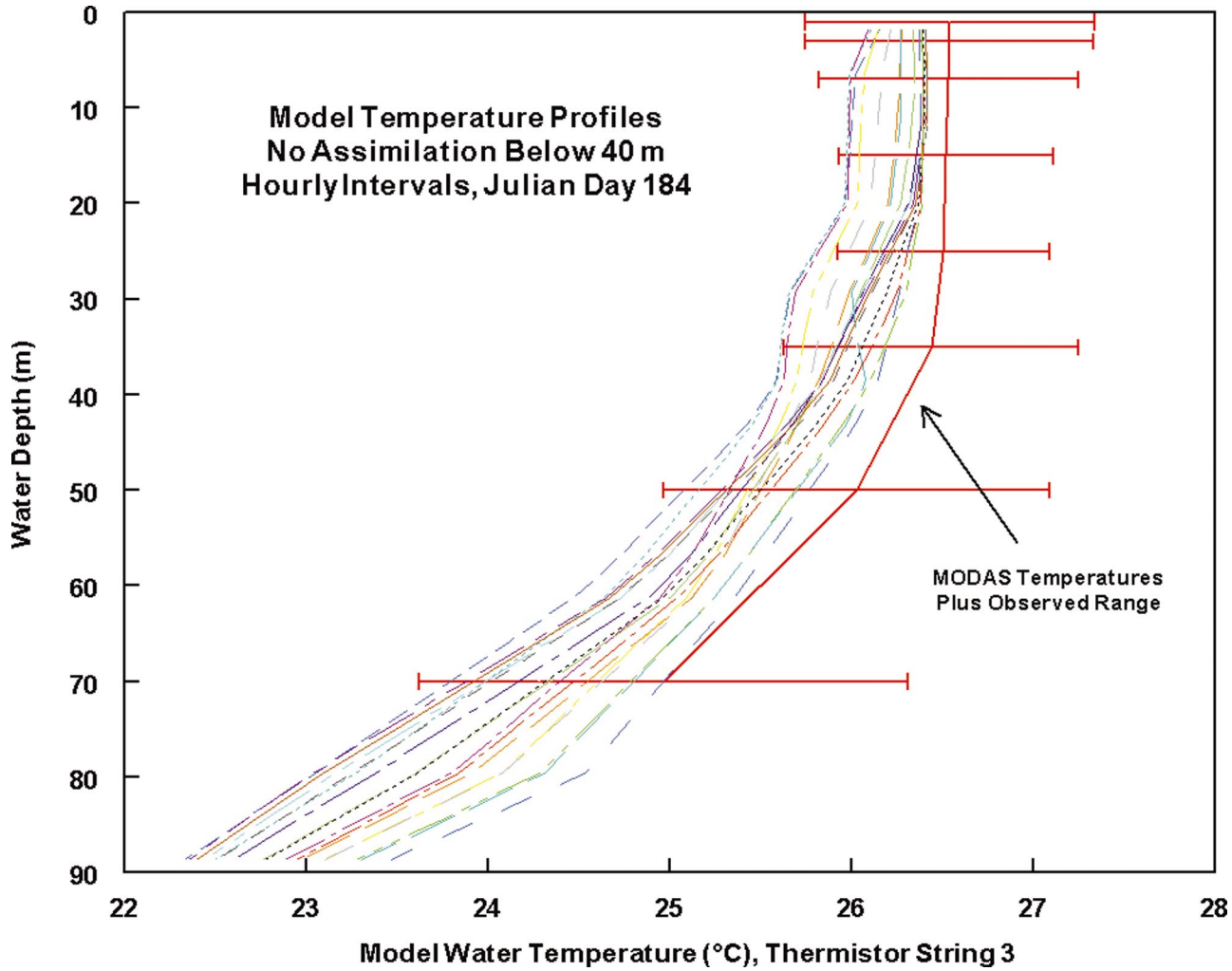


FIG. 11. Model-predicted hourly temperature profiles for 3 July 2003 at the thermistor site shown in Fig. 3. The MODAS temperature profiles for that date and location are also shown (solid red), along with the observed ranges of temperatures (horizontal solid red) determined from the 178 BTs, centered on the MODAS temperatures.

Comparing Figs. 6 and 9, we would conclude that assimilating the one ray path had both positive and negative impacts. On the positive side, the acoustic information reduced the number of occurrences in which the model-predicted water temperatures were too warm. However, the assimilation failed to eliminate the warm bias of the model predictions near the ocean surface and actually accentuated the cool bias of the model in the lower parts of the water column.

The warm bias of the model near the ocean surface was a persistent feature for 30 June–3 July 2003. Using this fact, we modified our process such that assimilation occurred only over the top 40 m of the ocean model. No adjustments based on the acoustic data were made below 40 m. The resulting temperature profiles are shown in Fig. 11, and these can be compared to the profiles in Figs. 6 and 9. We see that model-predicted near-surface temperatures better reflect the observed temperatures, both in magnitude and variability. The lower layers of the water column were still too cool, but the assimilation process did not accentuate this bias in this test.

VII. SUMMARY AND CONCLUSIONS

We have presented a model-oriented acoustic inversion and assimilation technique for arrival time anomalies from bottom-mounted sources and receivers. In calculating an observed arrival time, we attempted to eliminate the impacts of Doppler shifts and scattering at the ocean surface by averaging over a number of pings. The reference arrival time between a source-receiver pair was calculated using an acoustic propagation model based on model-predicted SSPs, the specified locations and depths of the source and receiver, and the available bathymetry between the source and receiver.

The arrival time anomalies from only one acoustic path were used in a test of this technology at PMRF in Hawaii. When assimilating over all model grid cells, the acoustic information helped in reducing the number of occurrences in which the model-predicted water temperatures were too warm. However, the assimilation technology for this case introduced a distinct cool bias in the lower levels of the water column (Fig. 9). A second test used the fact that the model near-surface temperatures at the location of the observations were consistently too warm, so assimilation occurred only over the top 40 m of the water column. In this case, knowledge of the model bias allowed the technology to result in better near-surface predictions without enhancing the bias in the lower layers of the water column.

In the first test, the changes in the character of the pro-

files (Figs. 6 and 9) are readily explained by the values calculated for the Kalman gain matrix **K**. When dealing with only one ray path, there is a single element of the matrix **K** for each model grid cell. This is multiplied times the corresponding element of $\mathbf{z}_R - \mathbf{H}\mathbf{x}^{\mathbf{F}}$ (which itself will be a scalar for a single ray path). Thus, a larger value of the element of **K** for a grid cell results in a greater modification in the model-predicted water temperature in that grid cell during the PSAS assimilation process. An analysis of **K** showed that larger values were associated with water depths from 80 to 300 m. As a result, the assimilation of travel time anomalies had the greatest impact on the lower levels of the model grid cells. This differential impact is quite evident when we consider Fig. 10.

Because of the smaller values of the Kalman gain, the surface layers were not cooled fast enough before the lower layers were cooled so much that the model-predicted arrival time anomaly matches the observed arrival time anomaly. The result is the overcooling of temperatures in lower levels of the model (the more prominent cold bias in Fig. 9 when compared to Fig. 6) and the continuing warm surface bias for the model results.

Since the PSAS assimilation scheme is critically dependent on the Kalman gain matrix **K**, we have attempted to make reasonable estimates of the model and observation error covariance matrices. For $\mathbf{P_F}$, the scaled error approach [Eq. (13)] appears to work better than $\mathbf{P_F}$ calculated using the model covariances [Eq. (12)]. For **R**, the use of the standard deviation of the observed arrival times provides a practical means for determining the component $\mathbf{e_z}$. The results using our estimates of $\mathbf{P_F}$ and **R** appear to provide a good assimilation methodology of the observed travel time anomalies (Fig. 8).

It is obvious that additional work is needed in applying modifications to individual model grid cells based on travel time information from bottom-mounted hydrophones. Utilizing higher frequencies means sampling conditions throughout the water column over relatively short distances (3–5 km). This is advantageous in that multiple surface bounce ray paths from bottom-mounted hydrophones are relatively stable and are integrating over horizontal scales across which *T*, *S*, and **U** tend not to vary by large amounts. On the other hand, there is the disadvantage that these ray paths may end up reflecting water conditions that can vary significantly in the vertical. In the first case we presented, our double-surface bounce acoustic arrival time data and the calculated Kalman gain **K** results in a bias in the model predictions (Fig. 9) even though it is limiting the range of model solutions for water temperature. Work needs to be pursued to apply the modifications determined by the analysis fields to minimize the development of such biases. In our second test case (Fig. 11), we used knowledge of a model bias to limit the vertical extent of the assimilation process and, as a result, produce better model predictions.

In addition, we need to better understand the impacts of not discerning the exact path of an acoustic ray. There will always be inaccuracies in the latitudes, longitudes, and depths of a source and receiver as well as the bathymetry between the two. This leads to errors in the reference travel time t_R and the ΔL's (distances a ray travels through various model grid cells), both determined from the results of the acoustic propagation modeling. Bathymetric survey technology typically results in errors in depths of the order of meters. And errors in source, receiver, and bottom-bounce depths may or may not be cumulative. However, the latitudes and longitudes of hydrophones at depth (50+ m) obtained from a ship rolling on the open ocean are likely a greater source of errors for t_R and the ΔL's. It would not be unreasonable to expect such errors to be as large as tens of meters. Perhaps having multiple paths over numerous transects may mitigate such errors, but this has yet to be determined.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research and the National Defense Center of Excellence for Research in Ocean Sciences. The KauaiEx Group is Peter Stein, James K. Lewis, and Subramaniam Rajan (Scientific Solutions, Inc.), Michael B. Porter, Paul Hursky, Martin Siderius (SAIC), Mohsen Badiey (Univ. of Delaware), Jerald Caruthers (Univ. of Southern Mississippi), William S. Hodgkiss, Kaustubha Raghukumar (Scripps Inst. of Oceanography), Daniel Rouseff, Warren Fox (Univ. of Washington), Christian de Moustier, Brian Calder, Barbara J. Kraft (Univ. of New Hampshire), and Keyko McDonald (SPAWARSYSCEN).

[1] W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge U.P., New York, 1995).
[2] B. Cornuelle, "Ocean acoustic tomography: Integral data and ocean models," in *Modern Approaches to Data Assimilation in Ocean Modeling*, edited by P. Malanotte-Rizzoli (Elsevier Science, Amsterdam, 1996).
[3] B. Cornuelle, W. Munk, and P. Worcetser, "Ocean acoustic tomography from ships," J. Geophys. Res. **94**, 6232–6250 (1989).
[4] C. G. Walter, "Data Assimilation for Ocean Acoustic Tomography," Ph.D. thesis, Applied Physics Lab., Univ. of Washington, 1999.
[5] S. E. Cohn, A. da Silva, Jing Guo, M. Sienkiewicz, and D. Lamich, "Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System," Mon. Weather Rev. **126**(11), 2913–2926 (1998).
[6] A. F. Blumberg and G. L. Mellor, "A description of a three-dimensional coastal ocean circulation model," in *Three Dimensional Coastal Models, Coastal and Estuarine Sciences, 4*, edited by N. S. Heaps, Amer. Geophys. Union Geophysical Monograph Board (1987), pp. 1–16.
[7] V. Casulli and R. T. Cheng, "Semi-implicit finite difference methods for three-dimensional shallow water flow," Int. J. Numer. Methods Fluids **15**, 629–648 (1992).
[8] J. K. Lewis, Y. L. Hsu, and A. F. Blumberg, "Boundary forcing and a dual-mode calculation scheme for coastal tidal models using step-wise bathymetry," in *Estuarine and Coastal Modeling III: Proceedings of the 3rd International Conference*, edited by M. Spaulding *et al.* (ASCE, New York, 1994), pp. 422–431.
[9] W. H. F. Smith and D. T. Sandwell, "Global sea floor topography from satellite altimetry and ship depth soundings," Science **277**, 1956–1962 (1997).
[10] J. K. Lewis, I. Shulman, and A. F. Blumberg, "Assimilation of Doppler radar current data into ocean circulation models," Cont. Shelf Res. **18**, 541–559 (1998).
[11] G. D. Egbert, "Tidal data inversion: interpolation and inference," Prog. Oceanogr. **40**, 53–80 (1997).
[12] N. Booij, R. C. Ris, and L. H. Holthuijsen, "A third-generation wave model for coastal regions, Part 1, model description and validation," J. Geophys. Res. **104**(C4), 7649–7666 (1999).
[13] I. R. Young, *Wind Generated Ocean Waves*, edited by R. Bhattacharyya and M. E. McCormick, Ocean Engineering Series (Elsevier, Amsterdam, 1999).
[14] L. H. Kantha and C. A. Paulson, *Numerical Models of Oceans and Oce-*

anic Processes (Academic, New York, 2000).
[15] J. K. Lewis, ''A three-dimensional ocean circulation model with wave effects,'' in *Estuarine and Coastal Modeling V*, edited by M. Spaulding and A. F. Blumberg (ASCE, New York, 1998), pp. 584–600.
[16] I. Shulman, C.-R. Wu, J. K. Lewis, J. D. Paduan, L. K. Rosenfeld, J. C. Kindle, S. R. Ramp, and C. A. Collins, ''High resolution modeling and data assimilation in the Monterey Bay area,'' Cont. Shelf Res. **22**, 1129–1151 (2002).
[17] M. B. Porter and H. P. Bucker, ''Gaussian beam ray tracing for computing ocean acoustic field,'' J. Acoust. Soc. Am. **82**, 1349–1359 (1987).
[18] M. B. Porter, P. Hursky, M. Siderius, M. Badiey, J. Caruthers, W. S. Hodgkiss, K. Raghukumar, D. Rouseff, W. Fox, C. de Moustier, B. Calder, B. J. Kraft, K. McDonald, P. Stein, J. K. Lewis, and S. Rajan, ''The Kauai Experiment,'' in *High Frequency Ocean Acoustics*, edited by M. Porter, M. Siderius, and W. A. Kuperman, AIP Conf. Proc. (AIP, New York, 2004).

Prediction of underwater sound levels from rain and wind

Barry B. Ma,[a] Jeffrey A. Nystuen, and Ren-Chieh Lien
University of Washington, Applied Physics Laboratory, 1013 NE 40th Street, Seattle, Washington 98105-6698

(Received 27 September 2004; revised 24 March 2005; accepted 24 March 2005)

Wind and rain generated ambient sound from the ocean surface represents the background baseline of ocean noise. Understanding these ambient sounds under different conditions will facilitate other scientific studies. For example, measurement of the processes producing the sound, assessment of sonar performance, and helping to understand the influence of anthropogenic generated noise on marine mammals. About 90 buoy-months of ocean ambient sound data have been collected using Acoustic Rain Gauges in different open-ocean locations in the Tropical Pacific Ocean. Distinct ambient sound spectra for various rainfall rates and wind speeds are identified through a series of discrimination processes. Five divisions of the sound spectra associated with different sound generating mechanisms can be predicted using wind speed and rainfall rate as input variables. The ambient sound data collected from the Intertropical Convergence Zone are used to construct the prediction algorithms, and are tested on the data from the Western Pacific Warm Pool. This physically based semi-empirical model predicts the ambient sound spectra (0.5–50 kHz) at rainfall rates from 2–200 mm/h and wind speeds from 2 to 14 m/s.
[DOI: 10.1121/1.1910283]

PACS numbers: 43.30.Pc, 43.30.Nb [AIT] Pages: 3555–3565

I. INTRODUCTION

The ocean ambient sound received by an omnidirectional hydrophone can be categorized into signals from natural and non-natural sources. In the frequency range from 0.5 to 50 kHz, sounds generated from natural sources are mainly caused by the physical processes at the ocean surface. The wind generates breaking waves by applying stress on the ocean surface and these breaking waves subsequently produce acoustically active bubbles. The sound generated from these bubbles can be related to the wind speed through an empirical algorithm at 8 kHz (Vagle *et al.*, 1990), allowing quantitative measurement of wind speed. Because of the loud and distinctive acoustic signal of rainfall (Nystuen and Selsor, 1997; Nystuen *et al.*, 2000), ocean ambient sound also provides a means to detect and measure oceanic rainfall. Since ambient noise from natural sources serves the baseline of the sound level in the ocean, the study of these sounds is not only important to the ocean physicists studying rainfall and wind over the ocean, or for the usage of sonars, but also for biologists studying the impact of anthropogenically generated sound on marine mammals. Marine mammals are one of the main sources of biologic noise in the ocean, since they utilize sound for hunting, communication, and social activities. The impact of anthropogenic produced noise pollution on the marine ecosystem is of increasing interest.

Naturally generated sound signals can be distinguished from non-natural sources by comparing the known spectral and temporal characteristics of each sound source. Recently, about 90 buoy-months of ocean ambient sound data have been collected using Acoustic Rain Gauges (ARGs) at different open-ocean locations. Distinct ambient sound spectra for various rainfall rates and wind speeds are identified through a series of discrimination processes (Ma and Nystuen, in press). Individual contributions from rainfall and wind to the overall sound production are separated by inspecting the frequency and temporal characteristics of the sound signal. Several physical processes have been identified as responsible for the production and absorption of sound. These include (1) wind generated breaking waves producing sound through sound generated by individual bubbles (over 500 Hz) (Medwin and Beaky, 1989), and (2) bubbles trapped by raindrop splashes (Nystuen, 2001; Medwin *et al.*, 1992). These processes lead to the generation of sound with distinctive spectral characteristics. By developing empirical algorithms for each of these characteristics, the ocean ambient sound level can be predicted given wind and rainfall conditions.

II. BACKGROUND PHYSICS

The primary natural sources of ocean ambient sound from 1 to 50 kHz are wind generated breaking waves and the splashes from raindrops. The principal mechanism is the resonant acoustic ''ring'' from bubbles trapped underwater by the breaking waves or the raindrop splashes. Different bubble sizes ''ring'' at their respective resonant frequencies. The frequency is given by

$$f = \frac{1}{2\pi a}\sqrt{\frac{3\gamma P_0}{\rho_0}}. \tag{1}$$

where f is the resonance frequency, a is the bubble diameter, P_0 and ρ_0 are the local pressure and density, and γ is the ratio of specific heats for air ($\gamma = 1.4$), and is a function of bubble size (Medwin and Clay, 1998). The spectral shape of wind generated sound has a relatively uniform negative spectral slope from a peak near 500 Hz to over 20 kHz. In this frequency band, the sound generated is thought to be due to

[a]Electronic mail: binbing@u.washington.edu

TABLE I. Acoustic raindrop sizes and types of bubbles generated. The raindrop sizes are identified by different physical mechanisms associated with the drop splashes.

Drop size	Diameter (mm)	Sound	Frequency range (kHz)	Bubbles generated	Splash character
Tiny	<0.8	Silent		No	Gentle
Small	0.8–1.2	Loud bubble	13–25	Type I	Gentle with bubble every splash
Medium	1.2–2.0	Weak impact	1–30	No	Gentle, No bubbles
Large	2.0–3.5	Impact, Loud bubbles	1–35 2–35	Types II, III	Turbulent Irregular bubble entrainment
Very large	>3.5	Loud impact Loud bubbles	1–50 1–50	Types II, III	Turbulent Irregular bubble entrainment Penetrating jet

the oscillation of individual bubbles (Medwin and Beaky, 1989).

As a raindrop hits the ocean surface, it generates sound both from the impact and from bubbles trapped under the surface during the splash (Medwin *et al.*, 1992). The rainfall signal varies, depending on the drop size distribution in the rain. The mean acoustic energy per drop for four raindrop sizes is described in Nystuen (1996). These spectra feature a peak at the resonance frequency of the ''dominant'' bubble sizes created by each size category. Because the dominant bubble size is different for different raindrop sizes, the spectral peaks associated with different raindrop sizes are at different frequencies. Three types of bubble generating mechanisms from raindrops have been identified. Type I bubbles are generated from small raindrops (size 0.8–1.2 mm) that trap a bubble described as ''the pinch off of the top of the splash crater by surface tension and gravity forces'' (Pumphrey *et al.*, 1989; Medwin *et al.*, 1990; Nystuen, 1993). These bubbles have resonant frequencies from 13 to 25 kHz. Type II bubbles are created by the large and very large raindrops (diameter$>$2.2 mm) when a canopy is formed roughly 50 ms after the initial impact. The closing crater produces a downward turbulent jet which penetrates the crater bottom, trapping large bubbles and producing the dominant sound from 1 to 10 kHz (Medwin *et al.*, 1992). Type III bubbles are ''delayed microbubbles'' occurring more than 100 ms after the initial impact of the large and very large raindrops and are apparently produced during the reentry splashes of the corona droplets thrown upward by the initial impact (Nystuen and Medwin, 1995). The frequencies of the type III bubbles are typically 2–35 kHz. The acoustic raindrop sizes and their characteristics are summarized in Table I.

Drizzle (light rain) has a unique signal from 13 to 25 kHz because of 1-mm raindrops present in most drizzles, and the lack of large raindrops in some drizzles. However, the signal is highly sensitive to the wind speed. This can be explained as a suppression of the bubble creation mechanism from small raindrops (Nystuen, 1993). Heavier rainfall, containing larger drops as well as small drops, produces very loud sound levels across a wider frequency range (Nystuen *et al.*, 1993). Because the distribution of bubble sizes produced by breaking waves and raindrops are different, one is able to identify the sound source (wind, rain or drizzle) by listening for distinctive spectral characteristics.

III. DATA DESCRIPTION

A. Ocean ambient sound data

The ocean ambient sound data were collected using ARGs. These instruments consist of an ITC-8263 hydrophone, signal preamplifiers, and a recording computer (Tattletale-8). The nominal sensitivity of these instruments is -160 dB relative to 1 V/μPa and the preamplifier instrument noise has the equivalent oceanic background noise level of about 28 dB relative to 1 μPa^2/Hz. Bandpass filters are present to reduce saturation from low frequency sound (high pass at 300 Hz) and aliasing from above 50 kHz (low pass at 40 kHz). The hydrophone sensitivity also rolls off above its resonance frequency, about 40 kHz. A data collection sequence consists of four 1024 point time series collected at 100 kHz (10.24 ms each) separated by 5 s. Each time series is fast Fourier transformed to obtain a 512-point (0–50 kHz) power spectrum. These four spectra are averaged together and spectrally compressed to 64 frequency bins, with frequency resolution of 200 Hz from 100 to 3000 Hz and 1 kHz from 3 to 50 kHz. These spectra are evaluated individually to detect the acoustic signature of rainfall and then are recorded internally. The temporal sampling strategy is designed to allow the instrument to record data for up to one year and yet detect the relatively short time intervals of associated with rainfall. In order to achieve this goal, the ARG enters a low power mode ''sleep mode'' between each data sample. For these deployments, the ARGs ''sleep'' for 8–9 min and then sample the sound field. If ''rain'' is detected, the sampling rate changes to 1 min (or 3–4 min if ''drizzle'' is detected) and stays at the higher sampling rate until rain is no longer detected. For the rain or drizzle detected condition, four spectra are stored (separated by 5 s). For no-rain condition, one averaged spectrum is stored.

The ARGs have been deployed on the Tropical Atmosphere Ocean (TAO) project (McPhaden *et al.*, 1998) moorings since 1998 at different locations with depths from 20 to 98 m. The depths are chosen to remain in the mixed layer and away from near surface transient noises. Assuming a

TABLE II. The acoustic data sets used in this study.

	Deployment			
Location	ARG	Depth (m)	Deploy period	Note
0, 165E	"B"	20	1 March 2000 –20 July 2000	The wind estimate higher than anemometer by 0–2 m/s
0, 165E	"K"	50	1 March 2000 –20 July 2000	Wind speed conversion well matches with anemometer
0, 165E	"D"	98	23 July 2000 –20 November 2000	Successful
0, 165E	"A"	28	July 2001 –January 2002	Successful
0, 165E	"G"	98	July 2001 –November 2001	Hydrophone malfunction after November
8N, 95W	"G"	38	December, 1999 March, 2000	Not much precipitation in this period
8N, 95W	"L"	38	22 April 2000– 10 November 2000	Successful
10N, 95W	"E"	38	December, 1999 –April, 2000	Not much precipitation in this period
10N 95W	"I"	38	21 April 2000 –November, 2000	Strong local noises
10N, 95W	"C"	38	10 November 2000 16 April 2001	Acoustic data at first 100 days, then noise present in two 10–20 day periods
10N, 95W	"L"	38	April, 2001–31 March 2002	Successful
12N, 95W	"F"	38	December, 1999 –April, 2000	Not much precipitation in this period
12N, 95W	"H"	38	23 April 2000 2 December 2000	Periodical local noises present
12N, 95W	"D"	38	27 March 2001 –1 April 2002	Successful

uniformly distributed source, the signal contains sound from the desired natural quantity (rain), plus sound from wind and noises from other sources. About 90 buoy months of acoustic data are used from two climate regions, the Western Pacific Warm Pool (WPWP) and the Inter-tropical Convergence Zone (ITCZ). Although the ARG is focused on rainfall detection, the long-term acoustical spectra from the deployment sites are available for wind speed measurement and various noise budget studies. The acoustic data used in this study are given in Table II.

B. Ancillary data

Ancillary data collected on the mooring are used for intercomparison of the rainfall and wind speed signal from the acoustic records.

1. Wind speed

Wind measurements are made using an anemometer mounted on the TAO mooring at 4 m height. The data values are 10-min averaged wind vectors. The error estimate is ± 0.3 m/s or 3% of the wind speed, whichever is greater (Freitag *et al.*, 2001). By using the Coupled Ocean-Atmosphere Response Experiment (COARE) V2.5 flux algorithm (Fairall *et al.*, 1996), these values are converted to equivalent 10 m height wind speed. The 10-m winds are slightly stronger than 4-m wind values by factors 1.010–1.0217. This is smaller than the error estimate of the anemometer.

2. Rainfall

Precipitation measurements on TAO buoys are made using a self-siphoning rain gauges mounted 3.5 m above the ocean surface. The instruments have a 100-cm^2 (11.3-cm diam) catchment cylinder mounted on top of a fill tube. The measuring tube has a maximum capacity of 500 mL, equivalent to 50 mm of rainfall accumulation, after which it automatically drains via a siphon. Siphon events take about 30 s,

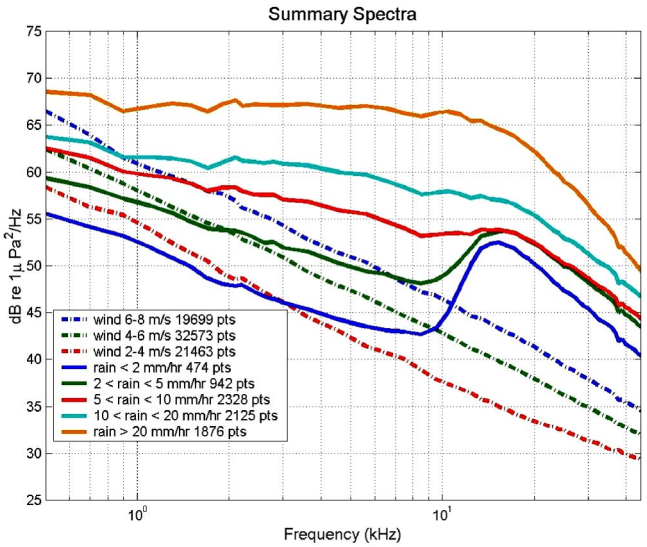


FIG. 1. Summary acoustic spectra on rain, drizzle, and wind speed.

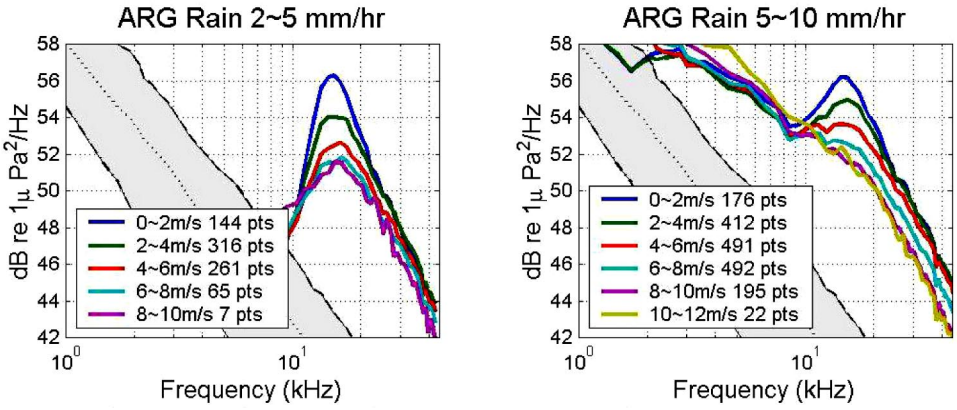


FIG. 2. Ambient sound spectra for 10 °N 95 °W from April 2001 to March 2002. The average sound spectra for rainfall rates of 2–5 and 5–10 mm/h are decomposed into various wind speeds. The shaded area represents the wind-only spectra where the dashed-dotted, dotted, and dashed lines are the average spectra for wind speeds at 2–4, 4–6, and 6–8 m/s, respectively. The numbers in the legend indicate how many spectra have been averaged to produce each spectrum.

and are typically identified by sharp declines in volume for two consecutive samples. In real-time processing, these events are ignored. The 1-min volume samples are stored on board the mooring while at sea, and are available for post-processing after recovery. In postdeployment processing, data associated with siphon events are flagged, typically removing 3 min worth of data centered on the event. Once the mooring is recovered, the 1-min accumulations are first flagged for obviously erroneous data. A 16-min Hanning filter is then applied to these data to generate smoothed 10-min accumulations. The estimated instrumental error for 10 min derived rainfall rates is 0.4 mm/h when the rain is present, and is 0.1 mm/h when there is no rain (Serra *et al.*, 2001).

IV. CHARACTERISTICS OF RAIN AND WIND GENERATED AMBIENT SOUND SPECTRA

Summary spectra of the rain and wind generated sounds are shown in Fig. 1 for all the available data from ITCZ and WPWP (Ma and Nystuen, in press). These spectra are grouped and averaged according to the collocated surface sensors' data, and provide a climatological overview of the ocean ambient sound. There are some distinctive features of sound generated by wind and rain. The wind-only spectra are shown for the wind speeds at 2–4, 4–6, and 6–8 m/s. These spectra (dash lines) have a constant negative slope from 1 to 50 kHz, with sound level decreasing at about 15.7 dB per decade. The rainfall spectra (solid lines) are averaged from the rainfall rate 2–5, 5–10, 10–20, and above 20 mm/h, respectively. Light rain (<10 mm/h) has a unique signal from 13 to 25 kHz, due to the 1-mm raindrops present in most drizzles. Heavier rainfall containing larger drops produces very loud sound levels across a wider frequency range. In the frequency band from 1 to 10 kHz the spectrum level increases with rainfall rate, but for frequencies above 10 kHz there is less dependence on the rainfall rate.

A. Wind influence on light rain—drizzle (2–10 mm/h)

In Fig. 1, no wind dependence of the rain generated sound is shown. To examine the wind influence on the rainfall generated sound, the surface anemometer data are used to further classify rainfall spectra. Figure 2 shows the spectra for rainfall rate 2–5 and 5–10 mm/h decomposed into various wind speeds. These data were collected in the ITCZ from April 2001 to March 2002. For these lighter rainfall rates (rainfall rate 2–5 and 5–10 mm/h), the rain generated sound at 15 kHz is inversely proportional to the wind speed. Laboratory studies show the probability that an individual drop will produce a bubble decreases linearly from 100% at normal incidence to 10% for oblique incidence 20° from the vertical (Medwin *et al.*, 1990; Nystuen, 1993). Thus, this decrease in sound level as the wind speed increases is explained as the suppression of the bubble trapping mechanism for small raindrops as the wind causes the impact angle of the raindrop splash to become more and more oblique. On the other hand, the sound level generated by rain in the frequency band from 2 to 8 kHz is relatively invariant with wind speed, suggesting that the bubble trapping mechanism for large drops is not sensitive to angle of impact.

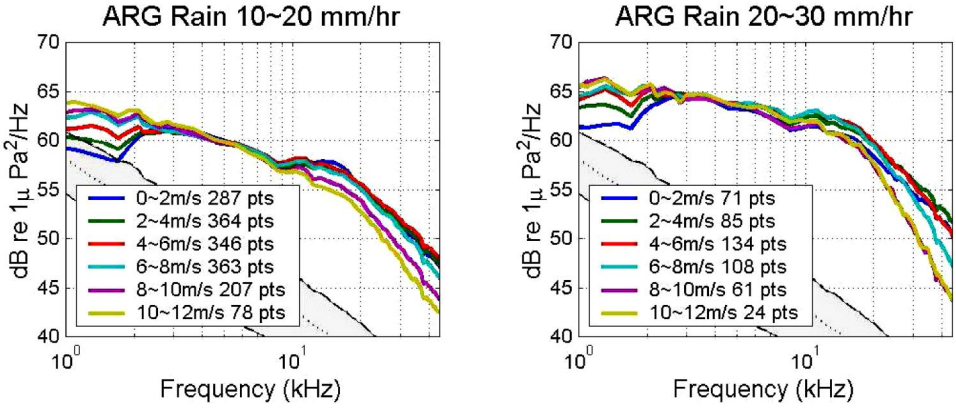


FIG. 3. The average sound spectra from rainfall rates of 10–20 and 20–30 mm/h are decomposed by wind speed. Shaded areas are the wind-only spectra (see Fig. 2 caption for details).

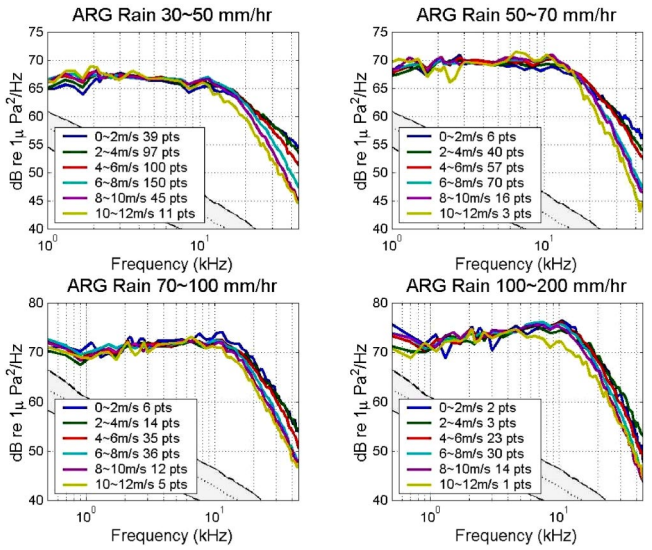


FIG. 4. The average sound spectra from rainfall rates of 30–50, 50–70, 70–100, and 100–200 mm/h are decomposed by wind speed. Shaded areas are the wind-only spectra (see Fig. 2 caption for details).

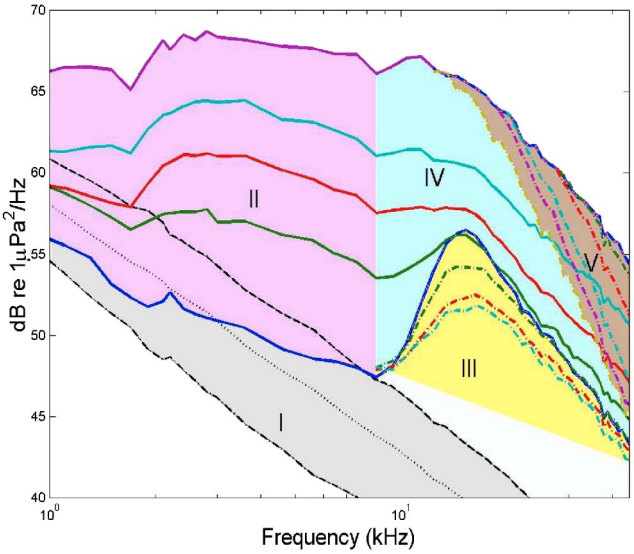


FIG. 5. The ocean ambient sound generated from wind and rain can be categorized into five sections with different sound producing and reduction mechanisms in the frequency band from 1 to 50 kHz: I. Wind-only. II. Large raindrops. III. Light rain (drizzle). IV. Both small and large raindrops. V. The masking effect due to layer of bubbles near the surface. The dash-dotted lines on sections III and V represent wind-dependent rainfall spectra.

B. Large raindrops signals in heavy rainfall (10–30 mm/h)

Figure 3 shows the sound spectra for the rainfall rate from 10 to 30 mm/h in various wind speed conditions. For the large and very large drops, the acoustic energy sharply drops off at below 2 kHz (Nystuen, 1996, Fig. 7). This is likely due to a limitation of the largest bubble sizes that can be produced by naturally occurring raindrop splashes. Figure 3 shows that below 2 kHz the spectra level increases as the wind speed increases. This is a summation effect, as the rainfall signal is relatively weak, while the wind signal is strongest in this frequency band and increases as wind speed increases. From 2 to 8 kHz the spectral level is relatively invariant to the wind speed and can be used to estimate rainfall rate. (Nystuen, 2001; Ma and Nystuen, in press). As the large raindrop population increases, the sound level from 10 to 40 kHz is also elevated and gradually dominates the sound generated from small raindrops. Thus the wind effect observed for light rain sound production can no longer be detected.

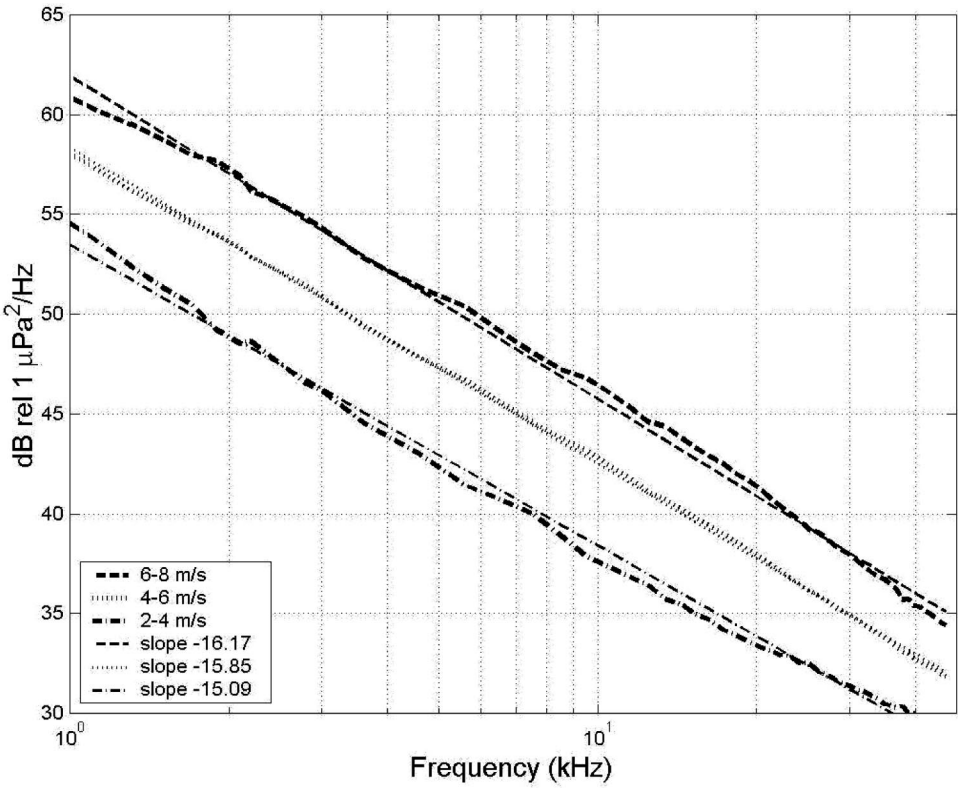


FIG. 6. The wind-only spectra and linear regressions for the wind speeds of 2–4, 4–6, and 6–8 m/s. This uniform slope −15.7 dB/decade is valid for the frequency band 1–50 kHz and wind speeds 2–14 m/s.

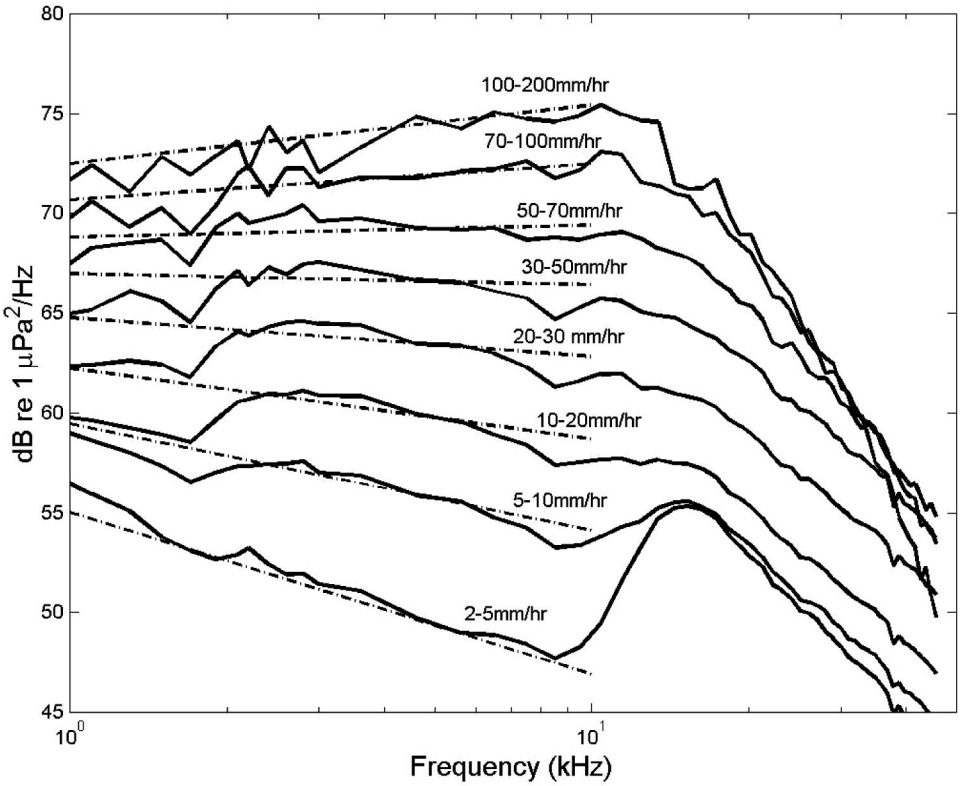


FIG. 7. The average rainfall spectra from discrete rainfall rate bins. The dashed lines are linear regressions for frequencies from 1 to 10 kHz. The slope of each linear regression line is given in Fig. 8.

C. Extremely heavy rainfall (30 mm/h above)

Once extremely heavy rainfall (>30 mm/h) is present, very large raindrops are usually present in the rain. These drops generate broadband sound (<1 kHz to >40 kHz) and little effect of wind is observed. In fact, the only wind dependence appears above 20 kHz where the sound level decreases as the wind speed increases (Fig. 4). Presumably, a thin layer of bubbles has formed near the surface and its effect is to attenuate subsequent sound produced on the surface both from wind and rain. This sound reduction effect has been observed in high wind speed conditions (Farmer and Lemon, 1984; Wille, 1984), and also during extremely heavy rainfall (>150 mm/h) (Nystuen *et al.*, 1993). The details of the interaction between these two types of bubble layer generating mechanisms are unknown. The influence of this bubble layer is observed for frequencies above 15 kHz, but not at lower frequencies. This is because the small bubbles that can be stirred down to form a ''layer'' absorb sound energy at high frequencies [Eq. (1)], while larger bubbles rise to the surface more quickly and consequently are not present to reduce sound levels at lower frequencies.

V. THE SEMI-EMPIRICAL ALGORITHMS

Five spectral sections associated with the different dominant sound generating and suppression mechanisms can be identified and used to form an empirical model to predict sound levels given the rainfall and wind speed conditions (Fig. 5). For section I, the wind-only spectra levels have the character of decreasing linearly as frequency increases. For section II, sound is produced by large raindrops and is invariant with respect to wind. For section III, between 8 and 25 kHz, sound is generated by small raindrops. This part of the spectrum is highly sensitive to the wind speed. For section IV, sound is produced by a combination of small and large raindrops. Finally for section V, there is a sound masking effect during high wind or extreme rainfall rates, due to a layer of bubble clouds which form just below the surface.

A. Section I: Wind generated noise from 1 to 50 kHz

Vagle *et al.* (1990) proposed an algorithm relating wind speed and sound pressure level (SPL) at 8 kHz,

$$U=[10^{(\mathrm{SPL}_{8\ \mathrm{kHz}}/20)}+104.5]/53.91, \tag{2}$$

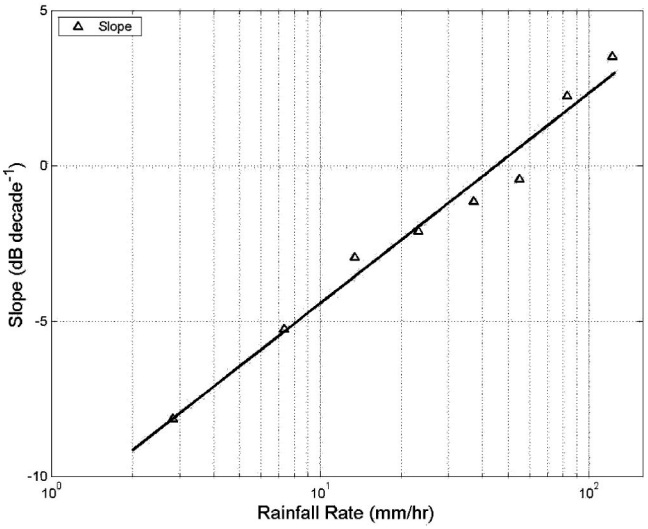


FIG. 8. The slope vs logarithmic rainfall rate (''Δ'' symbols) from the average rainfall spectrum for 1–10 kHz (Fig. 7). The rainfall rate is calculated using Eq. (5). The coefficients of the linear regression (solid line) are given in Eq. (8).

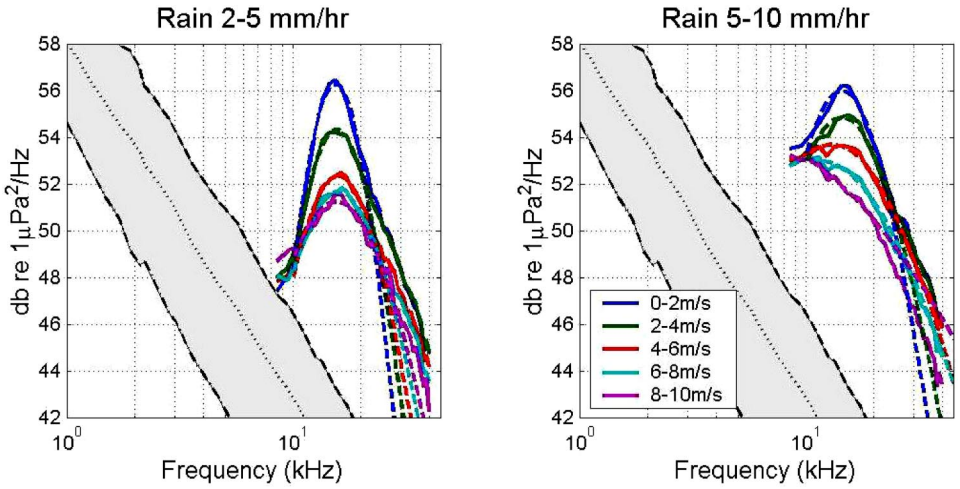


FIG. 9. The rainfall 2–5 mm/h (left) and 5–10 mm/h (right) in various wind speeds. The dashed lines are the least-squares fits using Eq. (9). The gray shaded area represents wind-only spectra for reference.

where U is wind speed at 10 m height (m/s), SPL is the sound pressure level in dB re 1 μPa^2/Hz. The data from the ITCZ and WPWP show that the wind-only spectra have a uniform slope from 1 to 50 kHz, and this uniform slope is held to at least 14 m/s. [Limited high wind speed data are recorded during the observation period, but it is known that high wind can induce a bubble layer near the surface which can mask the sound production at higher frequency (20 kHz plus) (Farmer and Lemon, 1984). Thus, the uniform slope would not be valid in this condition.] Linear regressions are applied to determine the slope of these "wind-only" spectra. This value is -15.7 dB/decade (Fig. 6). Using Vagle's wind speed algorithm at 8 kHz and slope deduced from the wind-only spectra, the empirical algorithm from wind-only sound can be written as follows:

$$\mathrm{SPL}_{\mathrm{wind\ 1-50\ kHz}}(f) = slope_{\mathrm{wind}}[\log(f) - \log(8)] + \mathrm{SPL}_{8\ \mathrm{kHz}}, \quad (3)$$

where $slope_{\mathrm{wind}} = -15.7$ dB/decade, and f is the frequency in kHz,

$$SPL_{8\ \mathrm{kHz}} = \log(U \cdot 53.91 - 104.5) \times 20, \quad (4)$$

where U is the wind speed in m/s.

B. Section II: Rainfall signal from 1 to 10 kHz

The sound for this part of spectrum is mainly generated from large raindrops (2.0–3.5 mm diameter). The population of 2–3.5 mm drops is highly correlated with rainfall rate and therefore can be used to quantify rainfall. The study of the data collected from the ITCZ (Ma and Nystuen, in press) shows the sound level at 5 kHz is proportional to the rainfall rate and relative invariant to the wind speed. The relation between the rainfall rate and sound pressure level at 5 kHz can be expressed as

$$\mathrm{dBR}/10 = (\mathrm{SPL}_{5\ \mathrm{kHz}} - 42.4)/15.4, \quad (5)$$

where $\mathrm{dBR} = 10\log_{10}(R)$, R is the rainfall rate (mm/h) and $\mathrm{SPL}_{5\,\mathrm{kHz}}$ is the sound level at 5 kHz in dB re 1 μPa^2 Hz^{-1}.

The average spectra for a series of discrete rainfall rates are shown in Fig. 7. The slopes from 1 to 10 kHz as a function of rainfall rate are shown in Fig. 8. It indicates that the slope increases linearly with logarithmic rainfall rate. Thus, the empirical algorithm for rain generated sound in this frequency band (1–10 kHz) is

$$\mathrm{SPL}_{1-10\ \mathrm{kHz}}(f) = [\log(f) - \log(5)] \cdot \mathrm{slope}_{\mathrm{rain}} + SPL_{5\ \mathrm{kHz}}(R), \quad (6)$$

$$\mathrm{SPL}_{5\ \mathrm{kHz}} = 15.4 \times \log(R) + 42.4, \quad (7)$$

$$\mathrm{slope}_{\mathrm{rain}} = 8.33 \times \log(R) - 14.3, \quad (8)$$

where the R is the rainfall rate in mm/h, f is the frequency in kHz, and $\mathrm{slope}_{\mathrm{rain}}$ is the slope for rainfall spectrum from 1 to 10 kHz. The coefficients for Eq. (8) are from the linear regression line shown in Fig. 8. Equation (5)–(8) are valid for rainfall rates from 2 to 200 mm/h in the frequency band 1–10 kHz.

TABLE III. The numerical values for three coefficients in Eq. (9).

Wind speed (m/s)	0–2 m/s	2–4 m/s	4–6 m/s	6–8 m/s	8–14 m/s
			Rainfall rates 2–5 mm/h		
A	224.9443	217.0343	209.0711	206.3746	204.8878
Lslope	1.8199	1.4904	1.2577	1.1173	1.0084
Hslope	−1.8993	−1.5843	−1.3486	−1.1938	−1.0421
			Rainfall rates 5–10 mm/h		
A	223.7316	219.2085	214.1092	210.2187	207.0281
Lslope	1.2810	1.0731	0.7853	0.7121	0.4557
Hslope	−1.2137	−1.0274	−0.7070	−0.5934	−0.2976

TABLE IV. (a) Log(A) for Eq. (10). (b) Log(lc) for Eq. (10). (c) hc for Eq. (10). (d) Islope for Eq. (10). (e) hslope for Eq. (10).

	Wind speed (m/s)	Rainfall rate (mm/h) 2–5	5–10	10–20	20–30	30–50	50–70	70–100	100–200
(a)	0–2	14.1957	13.8391	15.0699	12.3104	2.0785	1.9197	2.5930	1.9049
	2–4	2.1813	11.0307	15.0699	12.3104	2.0785	1.9197	2.5930	1.9049
	4–6	2.2325	2.4076	2.1078	2.2131	2.7322	1.8805	1.8933	1.9209
	6–8	12.5647	2.5482	2.3779	2.1635	1.9345	2.4050	1.9077	2.4000
	8–10	14.6970	2.2080	2.2181	2.2769	2.3827	1.9199	1.9263	2.0665
	10–12	⋯	2.2563	2.2181	2.2769	2.3827	1.9199	1.9263	2.0665
(b)	0–2	12.4299	12.0739	13.1704	10.4903	−0.1058	−0.6776	0.4561	−0.9591
	2–4	0.2169	9.2616	13.1704	10.4903	−0.1058	−0.6776	0.4561	−0.9591
	4–6	0.3045	0.5075	0.0294	0.1044	0.6885	−0.9660	−0.9480	−0.8317
	6–8	10.7961	0.6296	0.3132	−0.1925	−0.6335	−0.2254	−0.8116	−0.7305
	8–10	12.8245	−0.2055	0.0579	0.0215	0.0526	−0.7155	−0.6645	−0.5385
	10–12	⋯	−0.5628	0.0579	0.0215	0.0526	−0.7155	−0.6645	−0.5385
(c)	0–2	0.0454	0.0445	0.3923	0.0859	0.0413	0.0154	0.4321	0.0001
	2–4	0.0605	0.0736	0.3923	0.0859	0.0413	0.0154	0.4321	0.0001
	4–6	0.0852	0.1319	0.0824	0.1617	0.4083	0.0024	0.0010	0.0010
	6–8	0.1385	0.2774	0.3745	0.4374	0.0701	1.3341	0.0041	1.9518
	8–10	0.4586	0.9185	0.3838	0.5333	0.7530	0.0190	0.0075	0.2759
	10–12	⋯	1.7343	0.3838	0.5333	0.7530	0.0190	0.0075	0.2759
(d)	0–2	0.2711	0.2379	0.3681	0.2192	0.3897	0.8067	0.5762	1.5061
	2–4	0.3741	0.2437	0.3681	0.2192	0.3897	0.8067	0.5762	1.5061
	4–6	0.3535	0.3324	0.3881	0.4669	0.4840	1.3101	1.4103	1.4390
	6–8	0.2900	0.3923	0.5055	0.8256	1.1144	1.1233	1.3980	1.5609
	8–10	0.4673	0.7730	0.6331	0.8359	0.9257	1.3169	1.1899	1.3694
	10–12	⋯	0.9563	0.6331	0.8359	0.9257	1.3169	1.1899	1.3694
(e)	0–2	−5.2642	−3.8849	−1.0219	−1.6682	−2.1001	−2.6527	−1.0359	−9.3603
	2–4	−4.0761	−2.8379	−1.0219	−1.6682	−2.1001	−2.6527	−1.0359	−9.3603
	4–6	−3.1740	−1.7214	−1.5357	−1.2627	−1.0480	−5.2171	−6.6678	−6.4785
	6–8	−2.4861	−1.1529	−0.9296	−0.7487	−1.4518	−1.4518	−4.5984	−0.1006
	8–10	−1.5047	−0.4345	−0.8350	−0.8713	−0.8194	−2.8035	−3.8332	−0.4957
	10–12	⋯	−0.1646	−0.8350	−0.8713	−0.8194	−2.8035	−3.8332	−0.4957

C. Section III: Light rain (drizzle) 10–50 kHz (wind-dependent)

During light rain, the main sources of sound generated are from small raindrops (diameters 0.8–1.2 mm). Small raindrop populations are poorly correlated with total water content and the influence of wind affects the magnitude of the signal. The analysis from ITCZ data indicates that the sound generating mechanism from light rain exists at all rainfall rates above 2 mm/h. Below 2 mm/h, the signal is often contaminated by the wind. Above 10 mm/h, the drizzle signals are masked by sound generated from the large raindrops. A three-coefficient nonlinear equation is designed to reproduce the observed spectra levels in various wind speeds by assuming the wind speed is the only variant in this section of the spectrum:

$$\mathrm{SPL}_{10-50\ \mathrm{kHz}} = \frac{A}{\left(1+\left(\frac{f}{f_0}\right)^{\mathrm{lslope}}\right)\cdot\left(1+\left(\frac{f}{f_0}\right)^{\mathrm{hslope}}\right)}, \qquad (9)$$

where f is the frequency in kHz, f_0 is 15.5 kHz, A is amplitude, lslope is low-frequency slope and hslope is high-

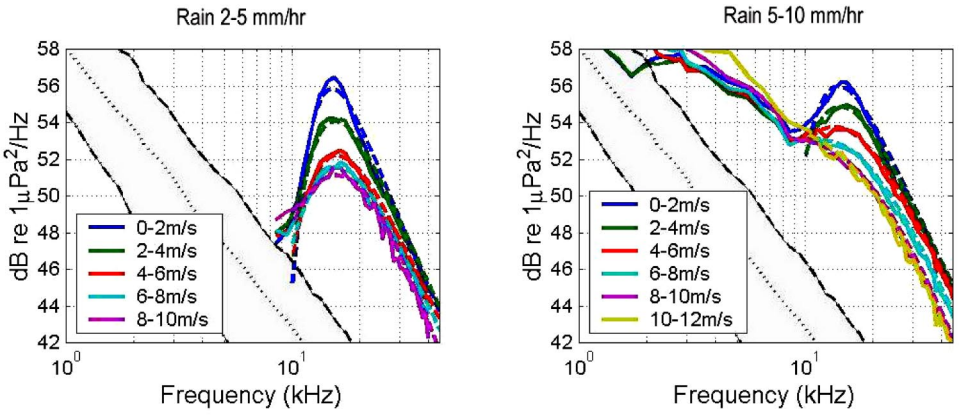


FIG. 10. The rainfall 2–5 mm/h (left) and 5–10 mm/h (right) in various wind speeds. The dashed lines are the least-squares fits using Eq. (10). The gray shaded area represents wind-only spectra for reference.

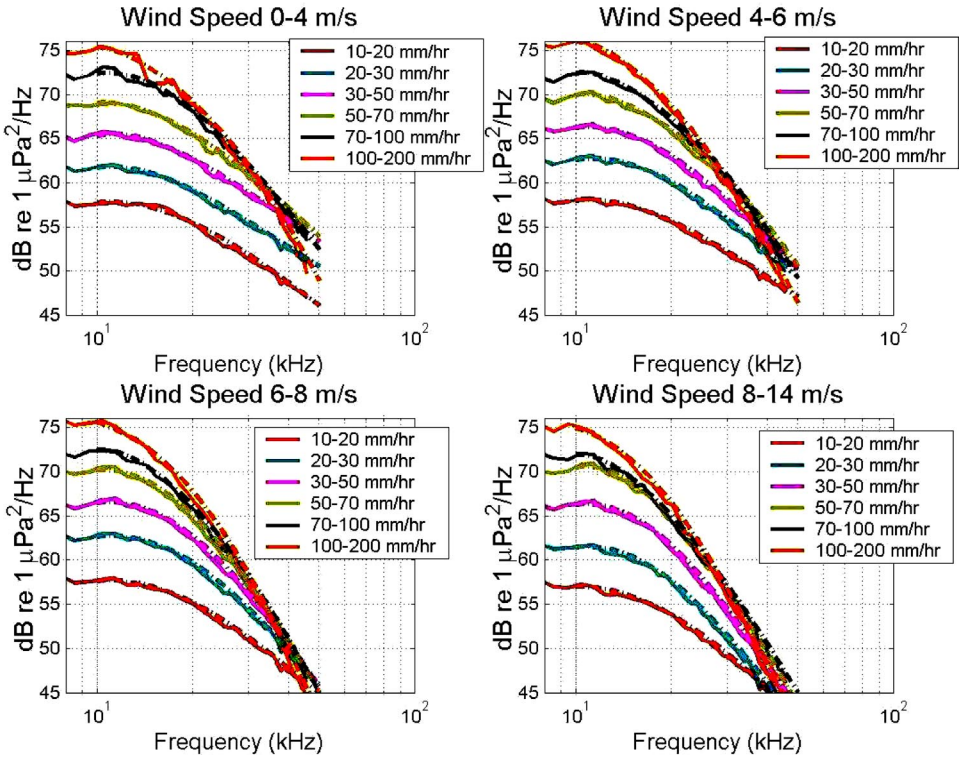


FIG. 11. The least-squares fits at frequency band 10 to 50 kHz for rainfall rates 10 mm/h and above. The fits are shown in dashed-dotted lines using Eq. (10). The numerical values of five coefficients are given in Tables IV (a)–(e).

frequency slope. The least-squares fits shown in Fig. 9 are applied to identify coefficients for each wind speed category. These coefficients are given in Table III. Equation (9) is a discrete algorithm, rather than the continuous algorithms developed for sections I and II. The absolute values of the high- and low-frequency slopes decrease as wind speed increases and the peak amplitudes are inversely proportional to the wind speed. In other words, the peak in the spectrum is suppressed as wind speed increases. Equation (9) is only good for frequencies up to 25 kHz. Above 25 kHz, the three-coefficient algorithm is insufficient to estimate observed spectrum levels (Fig. 9).

In order to produce a better result for frequency band above 25 kHz, a five-coefficient function is proposed:.

$$\mathrm{SPL}_{10-50\ \mathrm{kHz}} = \frac{A}{\left(1+\mathrm{lc}\left(\frac{f}{f_0}\right)^{\mathrm{lslope}}\right)\cdot\left(1+\mathrm{hc}\left(\frac{f}{f_0}\right)^{\mathrm{hslope}}\right)}, \tag{10}$$

where f is the frequency in kHz, f_0 is 15.5 kHz, and A, lc, hc, lslope, and hslope are five wind-dependent variables for this nonlinear equation. These numerical numbers are given in Table IV. The fitting result is shown in Fig. 10. We can try to assign physical meanings to the coefficients, but in fact these variables are statistically generated and may not have physical meaning, though the results show a better fit for frequencies >25 kHz.

D. Sections IV and V: Heavy to extremely heavy rainfall signals in the frequency band 10–50 kHz

During heavy to extremely heavy rainfalls (rainfall rates above 10 mm/h), the rainfall generated sounds at frequency band from 10 to 20 kHz are proportional to the rainfall rate and relative invariant to wind. For the high wind speed (10 m/s above) and/or extremely heavy rainfall (over 150 mm/h) conditions, a decrease in the sound levels in the frequency band from 20 to 50 kHz is observed. This is evidence of a subsurface bubble layer as previously described. The interaction mechanism of bubble formation from strong wind and/or heavy rain is not well understood. Consequently, the ambient sound data from ITCZ are sorted empirically according to the wind speeds and decomposed into various rainfall rates (Fig. 11). The least-squares fit is used to find the five coefficients for Eq. (10) (as rain-dependent variables this time). This gives a rain-dependent function to estimate the sound level at frequencies 10–50 kHz for four discrete wind speed bins (0–4, 4–6, 6–8, 8–12 m/s) when rainfall rate is greater than 10 mm/h. The values for these five coefficients in Eq. (10) are given in Table IV.

E. Summation

Finally, the wind-only and rain generated sound spectra are combined using power summation method:.

$$\mathrm{SPL}_{w,r} = 10\log_{10}(10^{(\mathrm{SPL}_{\mathrm{wind}}/10)} + 10^{(\mathrm{SPL}_{\mathrm{rain}}/10)}), \tag{11}$$

where $\mathrm{SPL}_{w,r}$ is the spectrum for wind and rain combined, $\mathrm{SPL}_{\mathrm{wind}}$ is the wind-only spectrum, and $\mathrm{SPL}_{\mathrm{rain}}$ is the rain spectrum. Note that a 28 dB relative 1 $\mu\mathrm{Pa}^2/\mathrm{Hz}$ laminate background noise is added to the wind-only spectrum when the wind speed is smaller than 2 m/s. This is due to the ARGs sampling dynamic limit and ocean background noise when neither wind nor rain is present. Equations (6)–(8) for the frequencies 1–10 kHz are continuous, but Eqs. (9) and (10) are discrete. In order to match the predicted sound levels at 10 kHz, discrete equivalent rainfall rates are used. The equivalent rainfall rate for each discrete rainfall rate bin is

TABLE V. The equivalent rainfall rate for each discrete data bin.

Rainfall rate bin (mm/h)	2–5	5–10	10–20	20–30	30–50	50–70	70–100	100–200
Equivalent rainfall rate (mm/h)	3.096	7.210	13.480	24.259	38.801	59.560	79.344	126.086

estimated by Eq. (5) for the frequency band 1–10 kHz (Table V).

VI. COMPARISON WITH OBSERVATIONS

A section of data during the summer rainy season from year 2000 at 0°, 165 °E is used for the modeling result comparison. The input data for the model are acquired from collocated anemometer and rain gauges. A subset of the original 1-min sampling interval of surface rainfall data is used after obvious errors (negative values and siphoning events) are removed. The wind speeds were recorded using a surface anemometer with a 10-min sampling interval and interpolated into 1-min intervals for model input. The acoustic data were collected using ARG at the depth of 50 m with 1-min sampling interval during the rain and 9 min during the no-rain periods. Figure 12 shows $3\frac{1}{2}$-h interval before and after a mesoscale convective system with maximum rainfall rate reaches 150 mm/h. There are no rainfall data available from surface rain gauge at 12:40 (Fig. 12, patched area). The new ambient sound model (Fig. 12, panel 3) mimics the observed sound field (Fig. 12, panel 2). Figure 13 is a 9-h interval before and after tropical rain storm with short and intense rainfalls during the storm. At 10:30, there is a low frequency noise in the acoustic record (Fig. 17, panel 2). And at 13:00, there is an unidentified high frequency noise. By comparing the model results with observations, the time and frequency of such non-nature sources can be extracted.

VII. CONCLUSIONS

Physical processes for five components of the naturally generated sound field are empirically modeled. This allows the prediction of ocean ambient sound from 1 to 50 kHz using two universal parameters, wind speed and rainfall rate. This model is a continuous function for frequencies 1–10 kHz and a discrete function for frequencies 10–50 kHz, and describes acoustic spectra of rainfall and wind with temporal resolutions of 1 min or less depending on the input data. There are also some limitations: (1) The algorithms are limited to a wind speed no greater than 14 m/s and a rainfall rate smaller than 200 mm/h. However this should cover most the natural variability of wind and rain. (2) For the wind-only sound prediction, the wind speed below 2 m/s does not produce wave breaking. Of course, sound levels are low under such conditions, but no correlation with ambient sound spectrum is expected. (3) The sound generated by the rainfall rates below 2 mm/h is often contaminated by the wind. The probability of rainfall detection is only about 30 % (Ma and Nystuen, in press), and thus sound predictions for rainfall

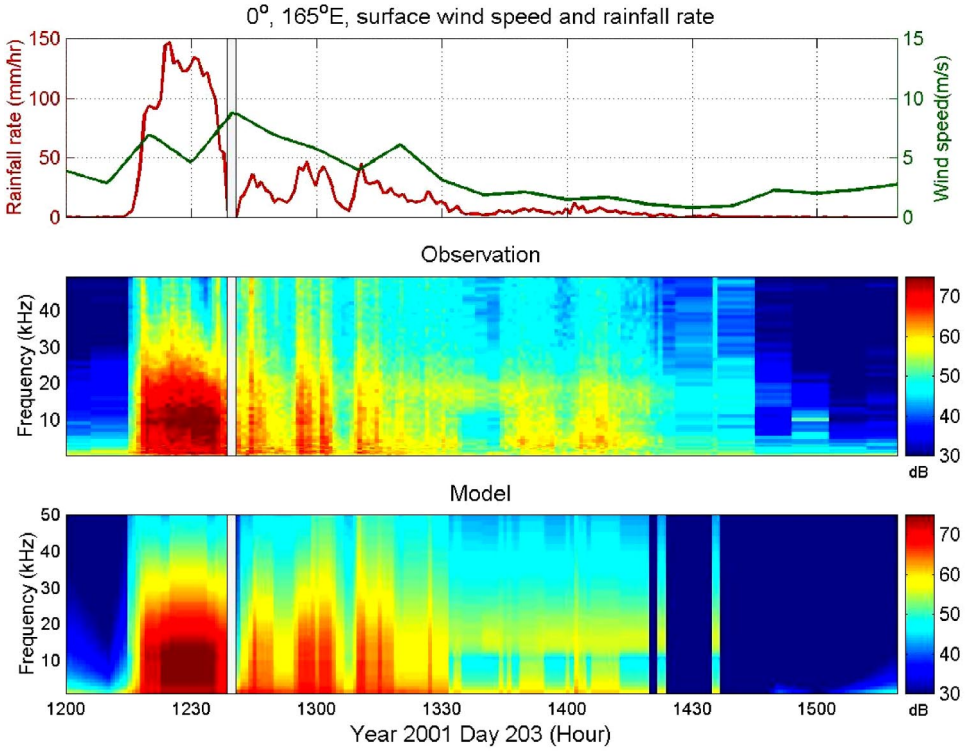


FIG. 12. Model results vs observation at 0°, 165 °E in Year 2000 Day 203. Top panel: The input wind speeds and rainfall rate data from surface instruments. Second panel: The observed sound filed (ARG). Third panel: The new ambient sound model. The gray patched area is the time when there was no surface rainfall rate data available.

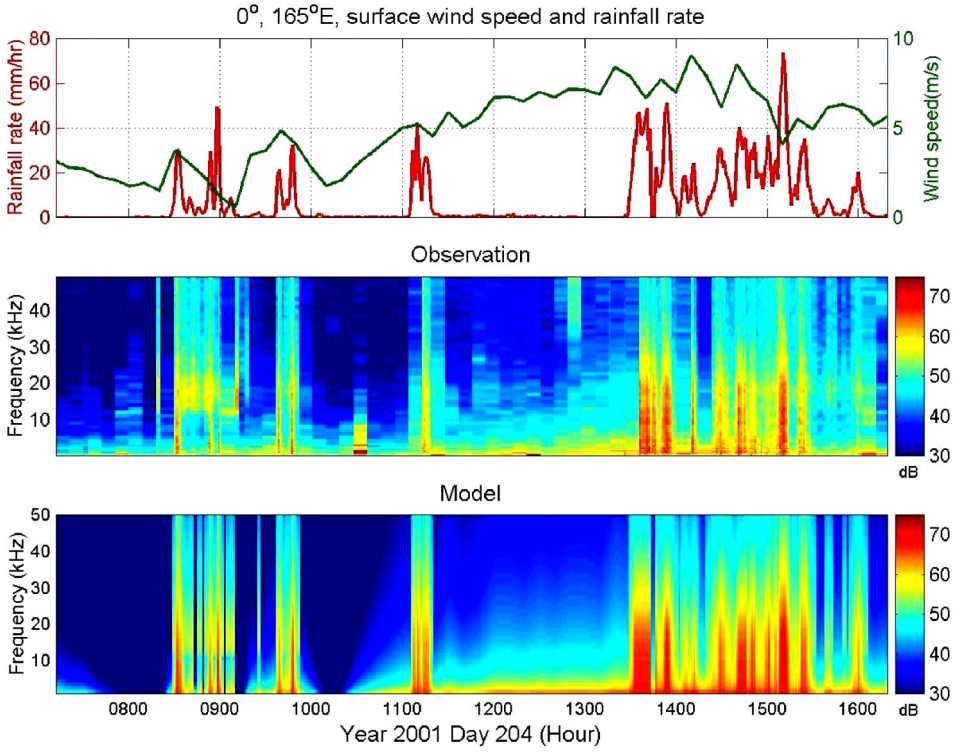


FIG. 13. Model results vs observation at 0°, 165 °E in Year 2000 Day 204. Top panel: The input wind speeds and rainfall rate data from surface instruments. Second panel: The observed sound filed (ARG). Third panel: The ambient sound model.

rate less than 2 mm/h are not modeled. (4) The sound levels at transition frequencies (1–2 and 8–10 kHz) may not be well estimated, since there is no clear understanding of the interaction between different sound production mechanisms. (5) A 28 dB $re\ 1\mu\text{Pa}^2/\text{Hz}$ background noise is added. This is due to ARG's dynamic limitation and/or the ocean background noise.

ACKNOWLEDGMENTS

The ARGs in this study were deployed and collected by the Climate Research Division of PMEL-NOAA, headed by Dr. Michael J. McPhaden. We wish to thank Dr. Eric D'Asaro, of APL-UW for his thoughtful comments and suggestions. Funding is from NSF Physical Oceanography, ONR Ocean Acoustics and the NASA TRMM Office.

Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B., and Young, G. S. (**1996**). "Bulk parameterization of air-sea fluxes for Tropical Ocean-Global Atmosphere Coupled-Ocean Atmosphere Response Experiment," J. Geophys. Res. **101**, 3747–3764.

Farmer, D. M., and Lemon, D. D. (**1984**). "The influence of bubbles on ambient noise in the ocean at high wind speeds," J. Phys. Oceanogr. **14**, 1762–1778.

Freitag, H. P., O'Haleck, M., Thomas, G. C., and McPhaden, M. J. (**2001**). "Calibration procedures and instrumental accuracies for ATLAS wind measurements," NOAA. Tech. Memo. OAR PMEL-**119**, NOAA/Pacific Marine Environmental Laboratory, Seattle, WA, 20 pp.

Ma, B. B., and Nystuen, J. A. (**2005**). "Passive acoustic detection and measurement of rainfall at Sea," J. Atmos. Ocean. Technol. (in press).

McPhaden, M. J., Busalacchi, A. J., Cheney, R., Donguy, J. R., Gage, K. S., Halpern, D., Ji, M., Julian, M., Meyers, G., Mitchum, G. T., Niiler, P. P., Picaut, J., Reynolds, R. W., Smith, N., and Takeuchi, K. (**1998**). "The Tropical Ocean-Global Atmosphere (TOGA) observing system: A decade of progress," J. Geophys. Res. **103**, 14169–14240.

Medwin, H., and Beaky, M. M. (**1989**). "Bubble sources of the Knudsen sea noise spectrum," J. Acoust. Soc. Am. **86**, 1124–1130.

Medwin, H., and Clay, C. S. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, New York), 712 pp.

Medwin, H., Kurgan, A., and Nystuen, J. A. (**1990**). "Impact and bubble sound from raindrops at normal and oblique incidence," J. Acoust. Soc. Am. **88**, 413–418.

Medwin, H., Nystuen, J. A., Jacobus, P. W., Snyder, D. E., and Ostwald, L. H. (**1992**). "The anatomy of underwater rain noise," J. Acoust. Soc. Am. **92**, 1613–1623.

Nystuen, J. A. (**1996**). "Acoustic rainfall analysis: Rainfall drop size distribution using the underwater sound field," J. Atmos. Ocean. Technol. **13**, 74–84.

Nystuen, J. A. (**1993**). "An explanation of the sound generated by light rain in the presence of wind," *Natural Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht), pp. 659–668.

Nystuen, J. A. (**2001**). "Listening to raindrops from underwater: An acoustic disdrometer," J. Atmos. Ocean. Technol. **18**, 1640–1657.

Nystuen, J. A., McGlothin, C. C., and Cook, M. S. (**1993**). "The underwater sound generated by heavy rainfall," J. Acoust. Soc. Am. **93**, 3169–3177.

Nystuen, J. A., and Medwin, H. (**1995**). "Underwater sound produced by rainfall: Secondary splashes of aerosols," J. Acoust. Soc. Am. **97**, 1606–1613.

Nystuen, J. A., McPhaden, M. J., and Freitag, H. P. (**2000**). "Surface measurements of precipitation from an ocean mooring: The acoustic log from the South China Sea," J. Appl. Meteorol. **39**, 2182–2197.

Nystuen, J. A., and Selsor, H. D. (**1997**). "Weather classification using passive acoustic drifters," J. Atmos. Ocean. Technol. **14**, 656–666.

Pumphrey, H. C., Crum, L. A., and Bjorno, L. (**1989**). "Underwater sound produced by individual drop impacts and rainfall," J. Acoust. Soc. Am. **85**, 1518–1526.

Serra, Y., Hearn, P. A., Freitag, H. P., and McPhaden, M. J. (**2001**). "ATLAS Self-siphoning rain gauge error estimates," J. Atmos. Ocean. Technol. **18**, 1989–2002.

Vagle, S., Large, W. G., and Farmer, D. M. (**1990**). "An evaluation of the WOTAN technique for inferring oceanic wind from underwater sound," J. Atmos. Ocean. Technol. **7**, 576–595.

Wille, P. C., and Geyer, D. (**1984**). "Measurements on the origin of the wind-dependent ambient noise variability in shallow water" J. Acoust. Soc. Am. **75**, 173–185.

Background noise cancellation for improved acoustic detection of manatee vocalizations

Zheng Yan
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Christopher Niezrecki[a)]
Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854

Diedrich O. Beusse
College of Veterinary Medicine, University of Florida, P.O. Box 100126, Gainesville, Florida 32610-0126

(Received 13 August 2004; revised 18 February 2005; accepted 16 March 2005)

The West Indian manatee (*Trichechus manatus latirostris*) has become endangered partly because of an increase in the number of collisions with boats. A device to alert boaters of the presence of manatees, so that a collision can be avoided, is desired. A practical implementation of the technology is dependent on the hydrophone spacing and range of detection. These parameters are primarily dependent on the manatee vocalization strength, the decay of the signal's strength with distance, and the background noise levels. An efficient method to extend the detection range by using background noise cancellation is proposed in this paper. An adaptive line enhancer (ALE) that can detect and track narrow band signals buried in broadband noise is implemented to cancel the background noise. The results indicate that the ALE algorithm can efficiently extract the manatee calls from the background noise. The improved signal-to-noise ratio of the signal can be used to extend the range of detection of manatee vocalizations and reduce the false alarm and missing detection rate in their natural habitat.
[DOI: 10.1121/1.1904385]

PACS numbers: 43.30.Sf, 43.60.Bf [WA] Pages: 3566–3573

I. INTRODUCTION

According to the United States Coast Guard, the number of registered boats in Florida has grown to over 900 000 as of 2001 (United States Coast Guard, 2002). The population of the West Indian manatee (*Trichechus manatus latirostris*) has also increased slightly in recent years, reaching an estimated population of 3276. Between 1995 and 2002 the percentage of mortalities of the West Indian manatee due to watercraft strikes has risen from 22% to 31% (Florida Department of Environmental Protection, Division of Marine Resources, 1996; Florida Fish and Wildlife Conservation Commission, 2002). This has lead to increased research into manatee avoidance technologies. A spatially fixed system to alert boaters of the presence of manatees is desired. Several methods to detect manatees have been proposed and include (1) a passive acoustic-based detection system (Herbert *et al.*, 2002; Mann *et al.*, 2002; Niezrecki *et al.*, 2003); (2) an above water infrared detection system (Keith, 2002); and (3) an underwater active sonar based system (Bowles, 2002). Each of these methods has their respective advantages and disadvantages. Within this paper, only the acoustic-based detection system is addressed. It is based on detecting the presence of manatees by using hydrophones that listen for manatee vocalizations. The frequency of manatee vocalizations, the source level of the vocalizations, the decay of the signal's strength with distance, and the amount of the ambient background noise will all affect the feasibility of an acoustic-based detection system. If the required hydrophone spacing is too small, the detection system will not be economically feasible. Within this study only a single hydrophone based system is being considered. The detection range for a single sensor system will ultimately determine the minimum hydrophone spacing for a complete system covering a broad area.

Schevill and Watkins were two of the first researchers to describe the vocalizations of the Florida manatee (Schevill and Watkins, 1965). Steel performed a detailed categorization of manatee vocalizations. Nine different categories of adult vocalizations were established. The fundamental frequency of a typical manatee call ranges from 2 to 5 kHz and the primary vocalization is a tonal sound that often has several harmonics. The second or third harmonic is often stronger than the fundamental frequency. Although some types of vocalizations are nonharmonic, the energy is dominated by one to three frequencies (Steel, 1982; Nowacek *et al.*, 2003). The vocalization rates of manatees during various activities have been observed by researchers. Apart from feeding, manatees vocalize approximately one to five times within a five-minute period. Bengston and Fitzgerald observed that vocalization rates are dependent on a manatee's behavior, with feeding and resting having the lowest vocalization rates and mating and cavorting being the highest. They also suggested that if manatee vocalizations are used for communicative and social purposes, then the vocalization rates might depend upon the number of manatees present (Bengston and Fitzgerald, 1985). Nowacek compared the vocalization rates

[a)]Electronic mail: Chris_Niezrecki@uml.edu

in Florida and Belize and found similar results. In some situations manatees can exhibit long periods of silence when no vocalizations are made (Nowacek *et al.*, 2003). The implementation of a passive acoustic detection system will certainly have to account for periods when a manatee is not vocalizing. Nowacek also estimated the mean received sound pressure levels of the peak frequency to be approximately 100 dB (re 1 μPa). The received values are recorded with the hydrophone at approximately 20 m from a group of 50 manatees. Position estimation techniques have also been used with hydrophones arrays to approximate the location of the vocalizing manatee. By using the estimated position and the received sound pressure levels, the mean source level of the manatee was approximated to be 112 dB @ 1 m (Phillips *et al.*, 2004). A comprehensive literature review on manatee vocalizations can be found in the work by Niezrecki *et al.* (2003).

Although there are many sources of underwater noise, there are two primary sources of background noise that typically corrupt a manatee call. One is boat noise and the other is snapping shrimp noise. Typical manatee vocalizations have a duration between 0.15 and 0.5 s. This feature makes the manatee call different in duration from the snapping shrimp noise, which typically lasts less than ten milliseconds. The spectrum of a typical manatee call, noises other than boat noise (called natural noise), and boat noise are shown in Figs. 1(a), 1(b), and 1(c), respectively. A manatee call is a narrow band signal compared to natural noise and boat noise. If the ratio between the signal strength of a manatee call and background noise is improved, the detection range can be extended. Conversely, the false alarm and missed detection rate will be higher when the manatee call is corrupted by high levels of background noise. Therefore, an efficient method to cancel the background noise is necessary in order to extend the range in which a manatee-warning device will be effective and to reduce the number of false alarms and missed calls.

For general adaptive noise cancellation algorithms, a reference signal that is not correlated with the desired signal in the primary input is required. A reference signal that is only correlated with the background noise and not the manatee calls is not available, for this particular problem. The underwater background noise is different from place to place, and is nonstationary. Therefore, we cannot use a hydrophone to get the noise reference signal for the adaptive filter. However, if we want to use an adaptive filter to reduce the background noise from the noisy manatee call, we must investigate another method to compensate for the absence of a noise reference input. The author's believe that an adaptive filtering approach is best suited to address this problem, as opposed to a Wiener filter (Gupta, 1984), because the manatee vocalizations and background noise are nonstationary.

In this paper an adaptive line enhancer that can detect and track narrow band signals in the broadband noise is implemented to cancel the background noise (Zeidler *et al.*, 1978). The noise reference input is replaced by a delayed version of the observed signal. The proper delay causes the broadband signal components in the reference input to be uncorrelated with those in the primary input and keeps a high correlation of the narrow band signal components in the reference input.

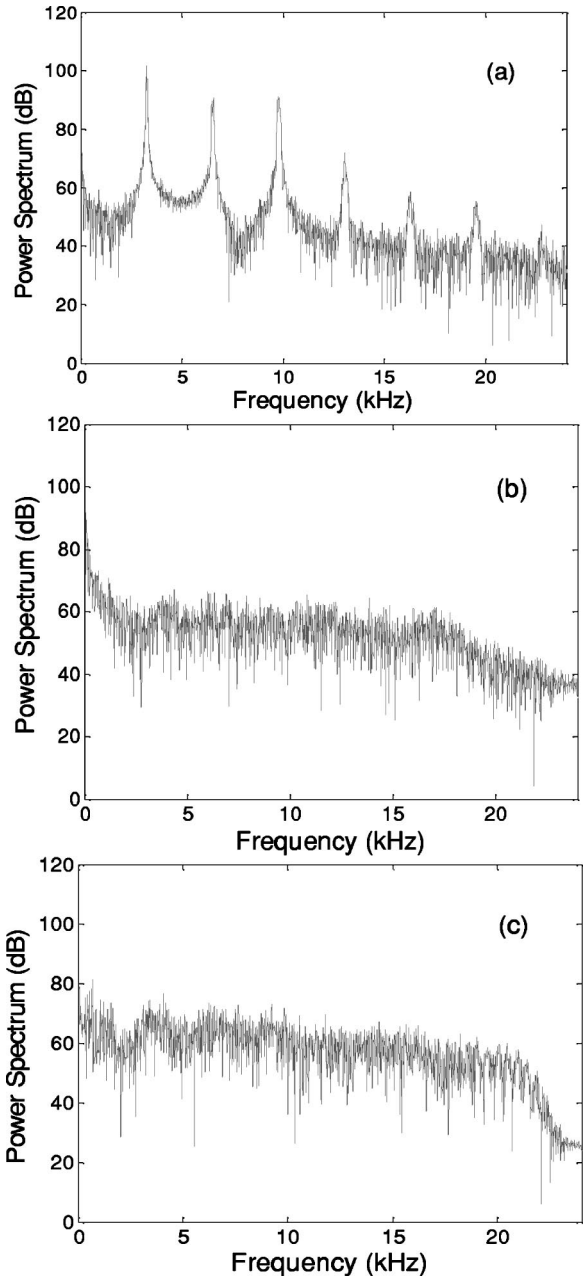


FIG. 1. (a) Power spectrum of a typical manatee vocalization; (b) power spectrum of typical natural noise; (c) power spectrum of typical boat noise.

This research represents the first effort to extend the range of manatee detection by using advanced signal processing techniques in order to extract manatee vocalizations that are buried in background noise. Specifically, the researchers are using active noise cancellation techniques for manatee signal enhancement. The theoretical development of the ALE algorithm is presented in Sec. II. The simulated results are shown in Sec. III, followed by the conclusions on the effectiveness of the proposed method.

II. THEORY DEVELOPMENT OF THE ALE

Widrow was the first to propose the adaptive line enhancer that is based on the Widrow-Hoff least-mean-square

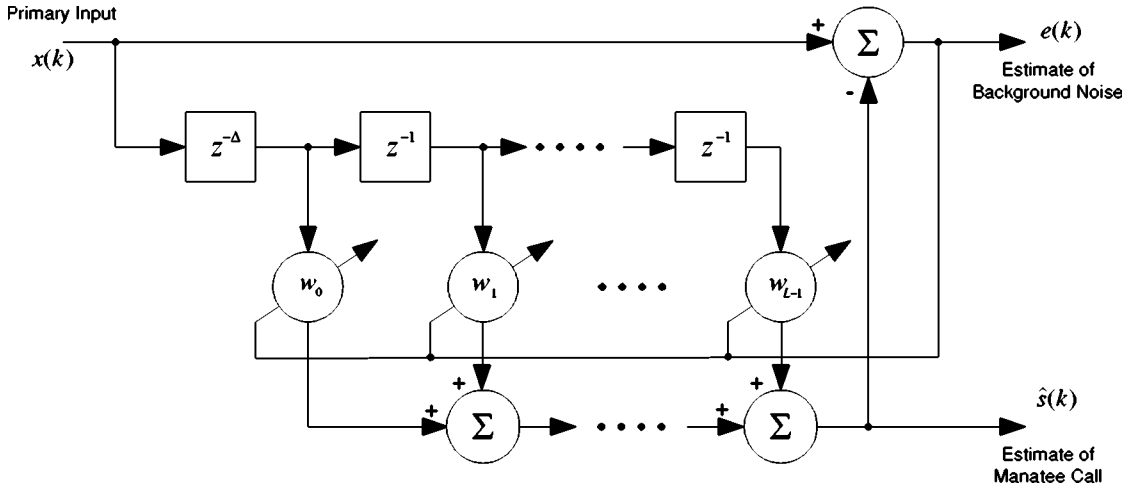


FIG. 2. Block diagram of an adaptive line enhancer (ALE).

(LMS) adaptive algorithm (Widrow, 1975). The ALE can be classified into two main categories: finite impulse response (FIR; an all zero filter) based and infinite impulse response-(IIR; a pole-zero filter) based algorithms. The FIR structure of ALE is shown in Fig. 2. The primary input $x(k)$ is assumed to be of the form

$$x(k)=s(k)+n(k), \tag{1}$$

where $s(k)$ is the sum of a number of narrow band components, and $n(k)$ is a zero-mean white Gaussian noise with power ν^2, which is independent of $s(k)$. For this problem, $s(k)$ represents the manatee vocalization, $n(k)$ the background noise, and $x(k)$ represents the hydrophone signal (primary input). The reference input is obtained by delaying the observed signal to give

$$x(k-\Delta)=s(k-\Delta)+n(k-\Delta). \tag{2}$$

The estimate of $s(k)$, which is $\hat{s}(k)$, is formed as

$$\hat{s}(k)=w_0x(k-\Delta)+w_1x(k-\Delta-1)+\cdots+w_{L-1}x(k-\Delta-L+1), \tag{3}$$

where the coefficients of adaptive filter, $\{w_i\}$, are updated by the error, $e(k)$, so as to minimize the cost function,

$$J=E[(x(k)-\hat{s}(k))^2]. \tag{4}$$

The delay must be of sufficient length to cause the broadband signal components in the reference input to become decorrelated from those in the primary input. The narrow band components will remain correlated with each other because of their periodic nature. The adaptive filter will form a transfer function equivalent to that of a narrow band filter centered at the frequency of the sinusoidal components. The noise component of the delayed reference input is rejected because it is uncorrelated with any components of the primary input. The phase difference between the sinusoid of

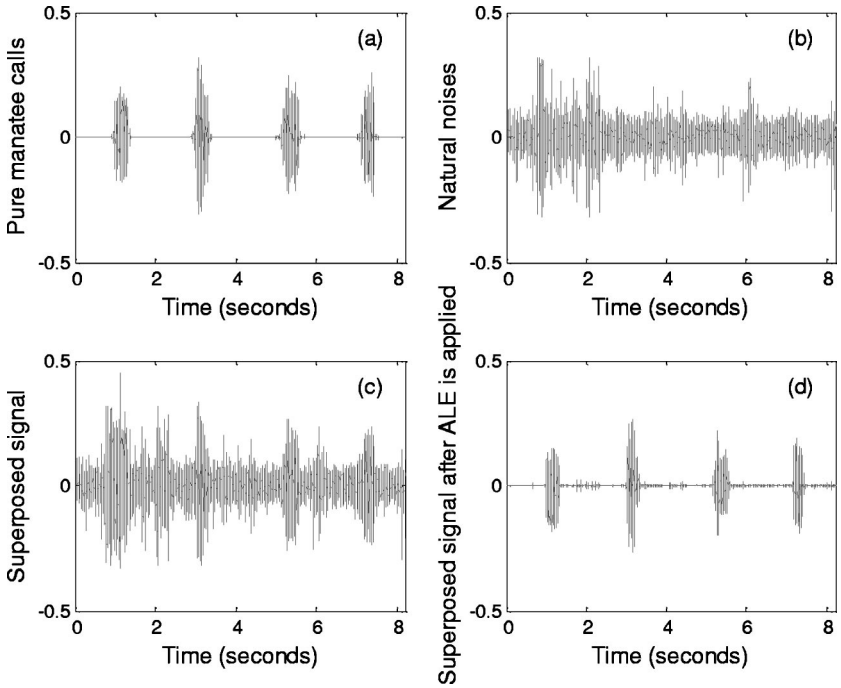


FIG. 3. (a) Pure manatee calls; (b) natural noise; (c) superposition of manatee calls and natural noise; (d) superposed signal after ALE is applied.

reference input and primary input is adjusted by the adaptive filter so that they cancel each other at the summation. The error sequence is fed back to adjust the weights of the adaptive filter according to the Widrow–Hoff LMS algorithm (Zeidler *et al.*, 1978),

$$w_i(k+1)=w_i(k)+2\mu e(k)x(k-i-\Delta), \quad i=0,1,...,L-1, \tag{5}$$

where μ is a constant. If the cancellation is ideal, the adaptive filter will cause $\hat{s}(k)$ to be a prediction of the current value of $s(k)$, and the error will be the broadband noise only.

The FIR structure of ALE is widely used because of its stability. However, it is computationally intensive when good accuracy is required. The order of ALE cannot be chosen to be too large because of the unavoidable misadjustment of the weight vector, which affects the minimum error the control system can reach. A larger order of ALE can narrow the filter pass band about the center frequency, thus improving the estimate of the signal amplitude for a given signal-to-noise ratio (SNR) of the input (Rickard and Zeidler, 1979). Therefore, there exists a tradeoff between the misadjustment of the weight vector and the estimate accuracy of the model. The FIR structure is always called suboptimal for modeling the underlying process because of its finite response. Although the IIR ALE can give an optimal solution, problems of instability and nonlinearity arise due to the feedback in IIR filters and nonlinearity can cause local minima in the error performance surface. The computation time of the IIR structure of ALE is reduced at the expense of some instability problems. Within this work, only the FIR structure of ALE is implemented.

As started earlier, a proper delay parameter should be selected to reject the broadband signal from the narrow band output of the LMS filter and extract the narrow band signal from the observed signal at the same time. Zeidler *et al.* were the first to study the effect of various choices of Δ on the steady-state response of the Δ-step predictor (Zeidler *et al.*, 1978). Reddy *et al.* and Egardt *et al.* derived the optimal delay parameter for one and two sinusoids when the signal-to-noise ratio is high, which gives not only a sharper spectral estimate but also an unbiased estimate of the sinusoid frequency. In order to find the optimum value of Δ for the ALE with a tapped-delay-line structure, they differentiated the average error variance with respect to Δ (Reddy *et al.*, 1981; Egardt *et al.*, 1983). Gupa and Yoganandam *et al.* derived the optimum value of Δ for one and two sinusoids under arbitrary SNR (Gupta, 1984; Yoganandam *et al.*, 1988). However, the optimal value of Δ for multiple sinusoids corrupted by broadband noise has not been derived yet. Deriving a solution for the optimal delay parameter for multiple sinusoids corrupted by broadband noise is complicated and not addressed in this paper.

If $p(j)$ represents the cross correlation between a signal and the delayed version of itself, and $r(j)$ represents the autocorrelation of that signal, then the relationship of $p(j)$ and $r(j)$ is given by

$$p(j)=r(\Delta+j), \tag{6}$$

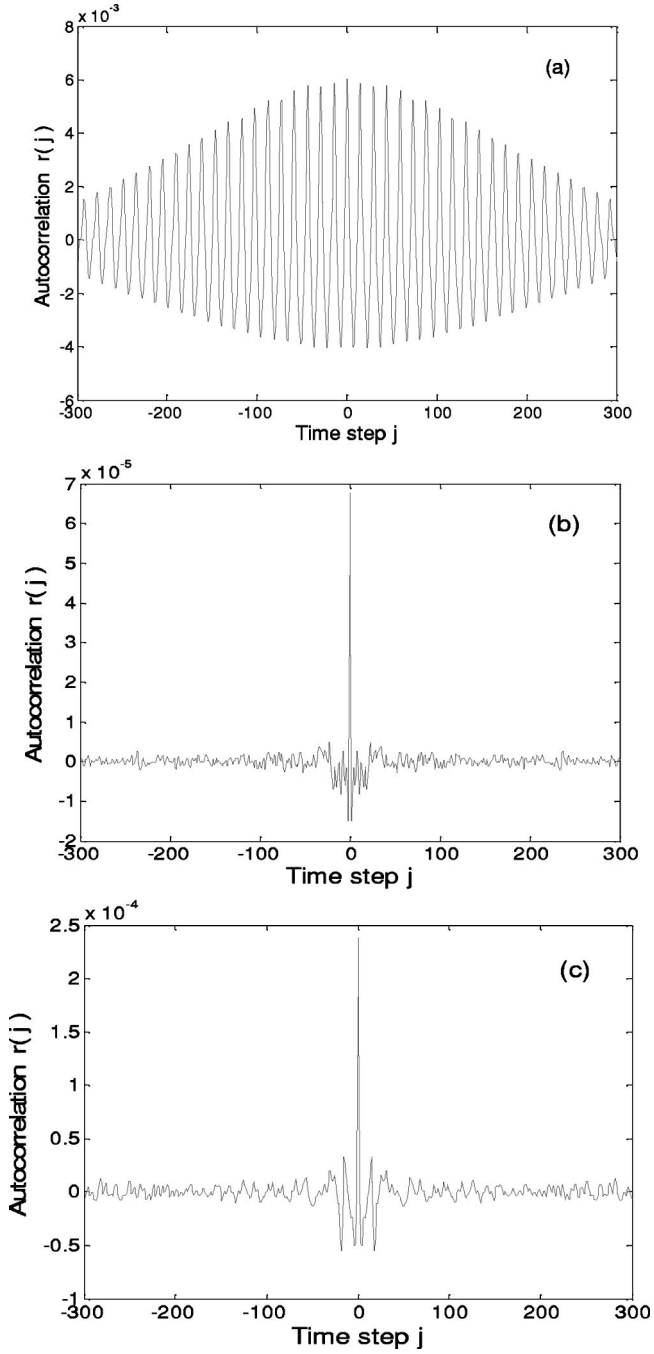


FIG. 4. (a) Autocorrelation of a pure manatee call; (b) autocorrelation of natural noise; (c) autocorrelation of boat noise.

where j is the lag time. For the ALE algorithm, $p(j)$ will be the cross correlation between the primary input $[x(k)]$ and the reference input [a delayed version of $x(k)$]. It is assumed that the manatee calls shown in Fig. 3(a) have negligible background noise. The data of a pure manatee call, the noises in the absence of the manatee call shown in Fig. 3(b), and the data of boat noise shown in Fig. 3(c) over the same period are used to calculate the autocorrelations of the manatee call, natural noise, and boat noise. All of these three autocorrelations are shown in Fig. 4. The results show that the envelope of the autocorrelation of the manatee call still remains high as lag j increases because of its periodic nature, but the autocorrelation of the natural noise and boat noise decreases

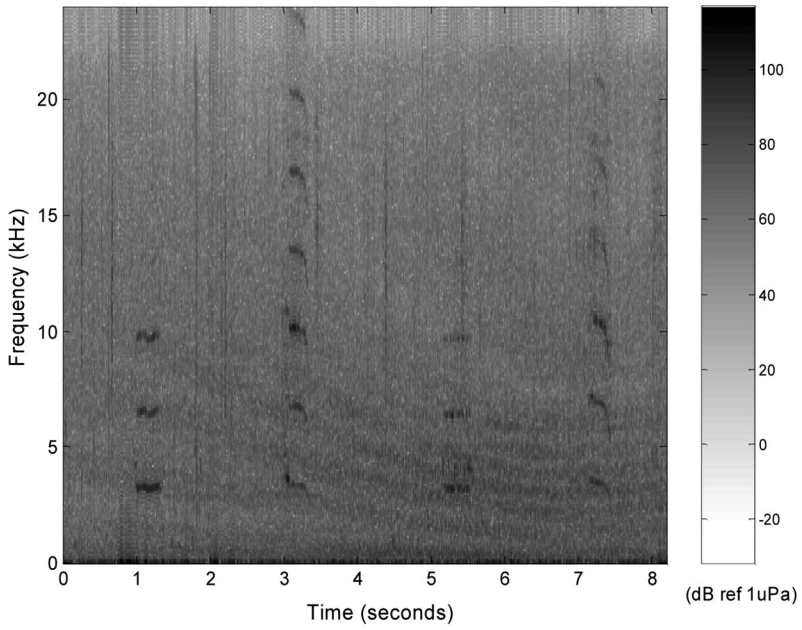


FIG. 5. The spectrogram of manatee calls corrupted by noises, including boat noise.

rapidly as lag j increases. Although the autocorrelation of the noises is small enough when $j=20$, simulation results showed that the delay is not effective in rejecting the boat noise from the narrow band output of ALE. Likewise, although the delay parameter, Δ, is large enough to decorrelate the white noise with $\Delta=1$, the simulations showed that the ALE with an optimal or near-optimal value of Δ yields a much better estimate of the sinusoid frequency (Reddy *et al.*, 1981; Egardt *et al.*, 1983). The delay parameter between 50 and 300 should be large enough to decorrelate the broadband noise in our case. However, the problem of optimal delay parameter for multiple sinusoids in broadband noise has not yet been solved. The delay parameter is set to 200 in these simulations.

The order of adaptive filter is another important factor that affects the performance of ALE. As stated earlier, increasing the filter length will narrow the filter passband about the center frequency, thus improving the estimate of signal amplitude for a given SNR of the input (Rickard and Zeidler, 1979). However, the misadjustment, defined as the dimensionless ratio of the average excess mean square error to the minimum mean square error, will go up with the increase of the order of the adaptive filter. The theoretical expression for misadjustment is given by (Widrow *et al.*, 1976)

$$M=\mu L\,\phi_{xx}(0), \tag{7}$$

where μ is the step size of the adaptive filter, L is the order of the adaptive filter, and ϕ_{xx} is the autocorrelation function for the input of ALE, $x(k)$, and given by

$$\phi_{xx}(j)=E[x(k)x(k+j)]. \tag{8}$$

Although the misadjustment level can be compensated by decreasing the step size μ, it will increase the convergence time of the adaptive filter. Sometimes a larger step size should be chosen to track a nonstationary signal. The SNR of output is limited due to the unavoidable misadjustment noise of the ALE. In general, the order is not suitable if it is too large. At the same time, a larger order filter requires a increased computation time. Nehorai and Malah derived the optimal order of the ALE for maximizing the SNR gain. When the adaptation time constant was specified and with the practical assumption that the N sinusoidal component have equal power and $\mu P_x L \ll 1$, the relationship between step size μ and the optimal order L_{opt} is given by

$$L_{\text{opt}} \cong [2N/(\mu P_x)]^{1/2}, \tag{9}$$

where N is the number of sinusoids and P_x is the total power of the input signal (Nehorai and Malah, 1980). All of these relationships can be used as guidelines for choosing the delay parameter, order, and step size of ALE. The order of the adaptive filter in our simulations is set to 70. The step size of the adaptive filter is a critical factor that affects the stability of the adaptive filter. Some algorithms that are aimed at speeding up convergence accomplish this at the expense of increased computational complexity, such as lattice structure, RLS algorithm, variable step size based on the correlation between the input signal and the error signal, and so on. However, the step size is set to be a constant between 0.05 and 0.5 in this work because of its low computational complexity. Therefore, it is important to carefully choose the delay parameter Δ, filter order, and the step size of the adaptive filter.

Since the fundamental frequency of manatee calls lies between 2 and 5 kHz, a high pass filter with a cutoff frequency of 1200 Hz is used to preprocess the noisy manatee call. This greatly reduces the energy of noises at low frequencies that may degrade the performance of the ALE. It

TABLE I. The performance of ALE for case I (dB).

	Call 1	Call 2	Call 3	Call 4
$\text{SNR}_{\text{before}}$	−1.14	8.54	5.87	6.51
$\text{SNR}_{\text{after}}$	37.93	33.79	30.35	32.25
G_{SNR}	39.07	25.25	24.48	25.74

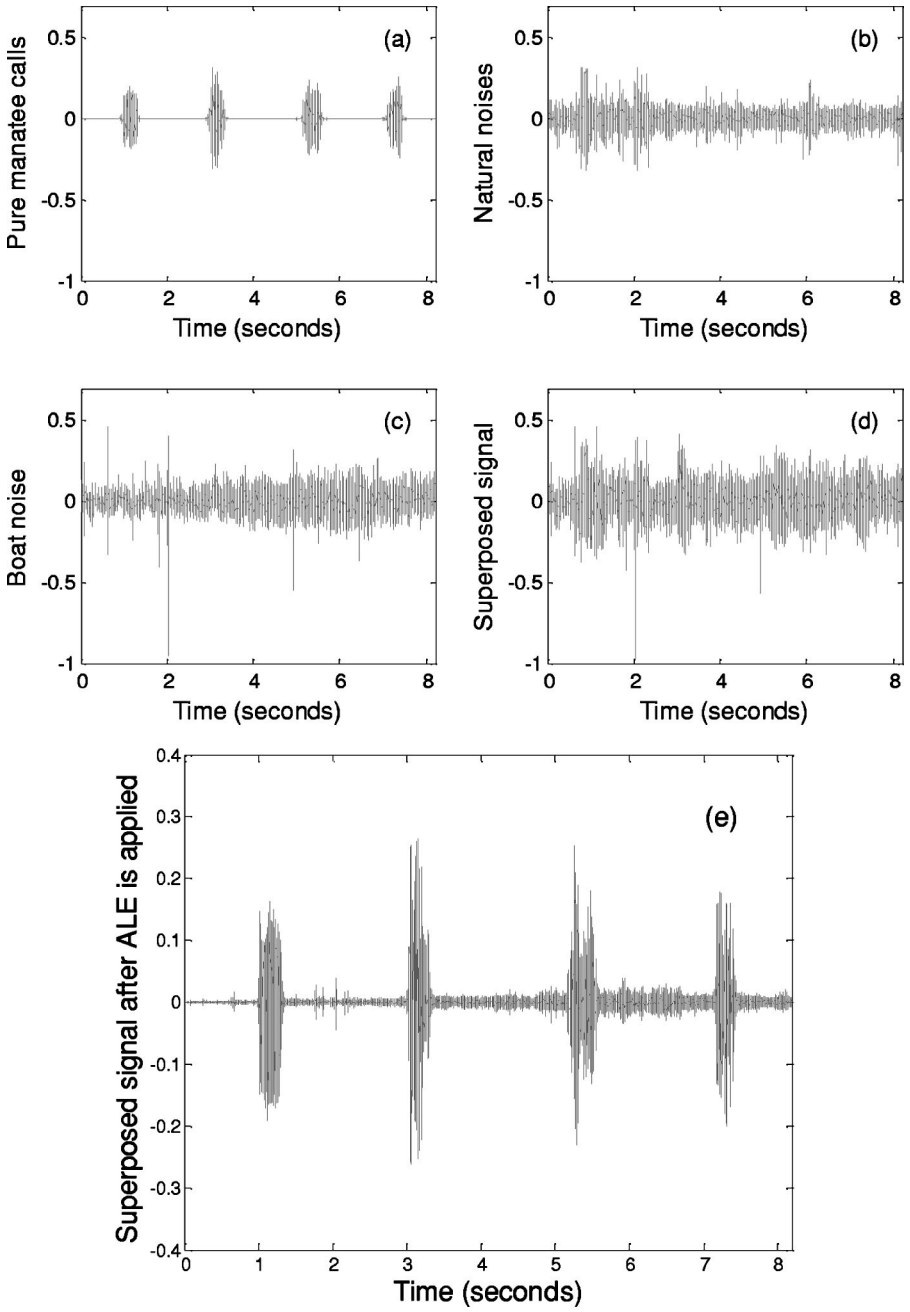


FIG. 6. (a) Pure manatee calls; (b) natural noise; (c) boat noise; (d) superposition of manatee calls, natural noise, and boat noise; (e) superposed signal after ALE is applied.

should be mentioned that a normalized LMS algorithm was also investigated and was shown to be not as effective in reducing the background noise compared to a LMS algorithm with a fixed step size. A normalized LMS algorithm is better suited for signals that are continuously measurable, unlike the manatee vocalizations.

III. SIMULATION RESULTS

O'Shea and the United States Geological Survey created an extensive library of manatee recording between 1981 and 1984 (O'Shea, 1981–1984). These recordings are used to quantify the performance of the ALE algorithm in this paper.

The measured signal corrupted by natural noise and boat noise represents the primary input to the adaptive filter. In practice, a manatee call signal may be corrupted by noises created by snapping shrimp, boats, rain, wind, fish, birds, marine mammals, and wave motion. The spectrogram of the manatee calls corrupted by noises is shown in Fig. 5. Two cases are simulated in this paper. The first case simulates a manatee call that is corrupted by natural noise. The second case includes the natural noise in case I but also includes boat noise. The synthesized data is constructed from actual measurements of manatee sounds, boat noise, and snapping shrimp noise that were measured separately. In case I, the pure manatee calls in the time domain and natural noise are shown in Figs. 3(a) and 3(b), respectively. The superposition of manatee calls and natural noise is shown in Fig. 3(c). The eight second hydrophone measurement contains four manatee vocalizations corrupted by background noise (snapping shrimp, wave motion, etc.). The manatee calls after ALE is applied is shown in Fig. 3(d). The four manatee calls are clearly more visible and much of the background noise has been filtered out. The performance of the ALE for case I is

TABLE II. The performance of ALE for case II (dB).

	Call 1	Call 2	Call 3	Call 4
SNR_{before}	−1.94	3.65	0.03	0.83
SNR_{after}	32.57	29.20	17.95	22.39
G_{SNR}	34.51	25.55	17.92	21.56

presented in Table I for the four calls. The signal-to-noise ratio gain, G_{SNR} is defined as

$$G_{SNR} = SNR_{after} - SNR_{before} \tag{10}$$

and represents signal-to-noise ratio improvement after ALE is applied, where SNR_{before} represents the signal-to-noise ratio of the manatee call before ALE is applied and SNR_{after} represents the signal-to-noise ratio of the manatee call after ALE is applied. The signal-to-noise ratio before ALE is applied is computed by taking the root mean square (rms) value of the time domain signals in the region where the manatee calls are present and dividing that value by the rms value in the corresponding region where the noise is present. The signal-to-noise ratio after ALE is applied is computed by same method. However, the noise is extracted from the near period of the calls in the time domain. The largest signal-to-noise ratio improvement is 39.07 dB.

In case II, boat noise is also included in the signal. The pure manatee call, other noises, high-level boat noise, and the sum of these three signals in the time domain are shown in Figs. 6(a), 6(b), 6(c), and 6(d), respectively. The superposed signal after ALE is applied is shown in Fig. 6(e). The performance of the ALE for case II is presented in Table II.

By inspecting Figs. 3 and 6, it is clear that the ALE has been able to clarify the four distinct manatee calls present in the signal. The instantaneous transfer function of the adaptive filter during the manatee call period is shown in Fig. 7. Although the transfer function cannot accurately form each peak of the manatee call due to the limited order of FIR, the magnitude of each peak of the transfer function is at least 15 dB larger than the surrounding frequencies. The adaptive filter provides approximately a 0 dB gain at the largest peak frequency of the manatee call. It is important to point out that the algorithm distorts the original manatee call; however, this is not critical since the signal is only used for detection and not reconstruction.

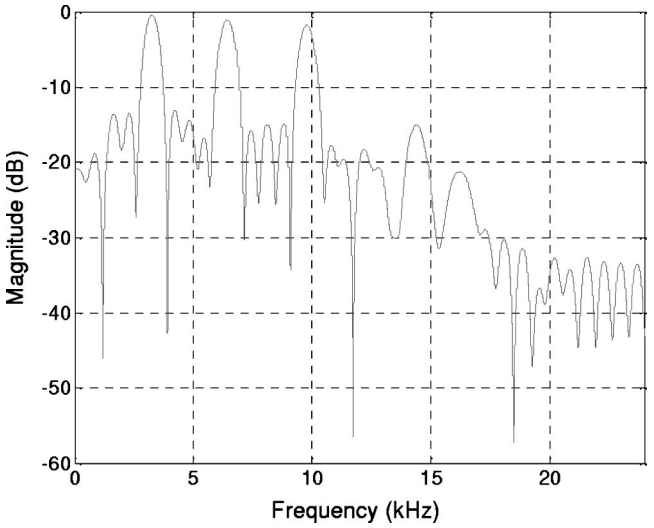


FIG. 7. The instantaneous transfer function of the adaptive filter during the manatee call period.

For the time domain data presented in Figs. 3 and 6, there are five discernable shrimp snaps over the 8.2 s of data. In the field, there are locations and durations in which the density of snapping shrimp noise will be higher (as many as tens of clicks per second). Within this work, the effectiveness of ALE has not been investigated for a higher density of snapping shrimp sounds. This issue is something that needs to be addressed in the future.

IV. CONCLUSIONS AND FUTURE WORK

The practical implementation of an acoustic-based manatee warning system is dependent on the minimum hydrophone spacing for the system. The required hydrophone spacing depends on the manatee vocalization strength, the decay of the acoustic signal's strength with distance, and the background noise levels. Within this paper, an adaptive line enhancer algorithm is used to reduce the underwater background noise in order to improve the signal-to-noise ratio of a set of manatee calls. The simulations showed that the ALE algorithm can efficiently discriminate the manatee calls from natural noise and high-level boat noise. The signal-to-noise ratio improvement ranges from 17.92 to 39.07 dB. The improved signal-to-noise ratio can be used to extend the detection range of manatee vocalizations, reduce the number of false alarms, and the number of missed calls. The results of this work may ultimately be used to realize a practical system that can warn boaters of the presence of manatees.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Sea Grant, Florida Fish and Wildlife Conservation Commission, and University of Florida Marine Mammal Program in supporting this research.

Bengston, J. L., and Fitzgerald, S. M. (**1985**). "Potential role of vocalizations in West Indian manatees," J. Mammal. **66**, 816–819.

Bowles, A. (**2002**). "Design for a manatee finder: sonar techniques to prevent manatee-vessel collisions," http://floridamarine.org/features/view_article.asp?id=14362.

Egardt, B., Kailath, T., and Reddy, V. U. (**1983**). "High-resolution spectral analysis using multi-step adaptive prediction," Circuits Syst. Signal Process. **2**, 422–443.

Florida Department of Environmental Protection, Division of Marine Resources (**1996**). Save the Manatee Trust Fund, Fiscal Year 1995–1996, Annual Report, Florida Marine Research Institute, 100 Eighth Avenue S.E., St. Petersburg, FL 33701-5095.

Florida Fish and Wildlife Conservation Commission (**2002**). Save the Manatee Trust Fund, Fiscal Year 2001–2002, Annual Report, Florida Fish and Wildlife Conservation Commission, 620 South Meridian Street, OESBPS, Tallahassee, FL 32399-1600.

Gupta, A. K. (**1984**). "Wiener filtering in PN spread-spectrum systems," *Proceedings of the IEEE MILCOM Conference*, Los Angeles, CA, October.

Herbert, T., Hitz, G., Mayo, C., Sermarini, C., Dobeck, G., Manning, B., Sandlin, M., Hansel, J., Bowden, T., and Artman, D. (**2002**). "Proof-of-concept for off the shelf technology to identify acoustic signature to detect presence of manatee(s)," http://floridamarine.org/features/view_article.asp?id=14362.

Keith, E. O. (**2002**). ''Boater manatee awareness system,'' http://floridamarine.org/features/view_article.asp?id=14362.
Mann, D., Nowacek, D., and Reynolds, J. III (**2002**). ''Passive acoustic detection of manatee sounds to alert boaters,'' http://floridamarine.org/features/view_article.asp?id=14362.
Nehorai, A., and Malah, D. (**1980**). ''On the stability and performance of the adaptive line enhancer,'' *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 5, April, pp. 478–481.
Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). ''Acoustic detection of manatee vocalizations,'' J. Acoust. Soc. Am. **114**, 1640–1647.
Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). ''Intraspecific and geographic variation of West Indian manatee,'' J. Acoust. Soc. Am. **114**, 66–69.
O'Shea, T. (**1981–1984**). ''Manatee vocalization—catalog of sounds,'' produced by Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL.
Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). ''Determination of West Indian manatee vocalization levels and rate,'' J. Acoust. Soc. Am. **115**, 422–428.
Reddy, V. U., Egardt, B., and Kailath, T. (**1981**). ''Optimized lattice-form adaptive line enhancer for a sinusoidal signal in broad-band noise,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-29**, 702–710.
Rickard, J. T., and Zeidler, J. R. (**1979**). ''Second-order output statistics of the adaptive line enhancer,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**, 31–39.
Schevill, W. E., and Watkins, W. A. (**1965**). ''Underwater calls of *Trichechus* (manatee),'' Nature (London) **205**, 373–374.
Steel, C. (**1982**). ''Vocalization patterns and corresponding behavior of the West Indian manatee (Trichechus manatus),'' Ph.D. dissertation, Florida Institute of Technology, Melbourne, FL.
United States Coast Guard (**2002**). ''Boating Statistics—2001,'' Commandant Publication P16754.15, 2100 Second Street SW, Washington, DC 20593-0001.
Widrow, B., Glover, Jr., J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, Jr., E., and Goodlin, R. C. (**1975**). ''Adaptive noise cancelling: principles and applications,'' Proc. IEEE **63**, 1692–1716.
Widrow, B., McCool, J., Larimore, M., and Johnson, Jr., C. (**1976**). ''Stationary and nonstationary learning characteristics of the LMS adaptive filter,'' Proc. IEEE **64**, 1151–1162.
Yoganandam, Y., Reddy, V. U., and Kailath, T. (**1988**). ''Performance analysis of the adaptive line enhancer for sinusoidal signals in broad-band noise,'' IEEE Trans. Acoust., Speech, Signal Process. **36**, 1749–1757.
Zeidler, J. R., Satorius, E., Chabries, D., and Wexler, H. (**1978**). ''Adaptive enhancement of multiple sinusoids in uncorrelated noise,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 240–254.

Comparison of multifrequency acoustic and *in situ* measurements of zooplankton abundances in Knight Inlet, British Columbia

Mark V. Trevorrow[a)]
Defence Research and Development Canada—Atlantic, PO Box 1012, Dartmouth, Nova Scotia, B2Y 3Z7, Canada

David L. Mackas
Institute of Ocean Sciences, 9860 W. Saanich Road, Sidney, British Columbia, V8L 4B2, Canada

Mark C. Benfield
Department of Oceanography and Coastal Sciences, Louisiana State University, Wetland Resources, Baton Rouge, Louisiana 70803

(Received 18 June 2004; revised 28 February 2005; accepted 29 March 2005)

An investigation of midwater zooplankton aggregations in a coastal fjord was conducted in November 2002. This study focused on quantitative comparisons between a calibrated, three-frequency (38, 120, and 200 kHz) vessel-based echo-sounder, a multinet towed zooplankton sampler (BIONESS), and a high-resolution underwater camera (ZOOVIS). Daytime layers of euphausiids and amphipods near 70–90-m depth were observed in lower parts of the inlet, especially concentrated by tidal flows around a sill. Quantitative backscatter measurements of euphausiids and amphipods, combined with *in situ* size and abundance estimates, and using an assumed tilt-angle distribution, were in agreement with averaged fluid-cylinder scattering models produced by Stanton and Chu [ICES J. Mar. Sci. **57**, 793–807, (2000)]. Acoustic measurements of physonect siphonophores in the upper inlet were found to have a strong 38-kHz scattering strength, in agreement with a damped bubble scattering model using a diameter of 0.4 mm. In relatively dense euphausiid layers, ZOOVIS abundance estimates were found to be a factor of 2 to 4 higher than the acoustic estimates, potentially due to deviations from assumed euphausiid orientation. Nocturnal near-surface euphausiid scattering exhibited a strong (15 dB) and rapid (seconds) sensitivity to vessel lights, interpreted as due to changing animal orientation. [DOI: 10.1121/1.1920087]

PACS numbers: 43.30.Sf, 43.30.Ft [KGF] Pages: 3574–3588

I. INTRODUCTION

High-frequency echo-sounders can be useful for assessing abundance and spatial distributions of meso-zooplankton such as crustaceans, pteropods, and siphonophores. Typical zooplankton echo-sounders, operating at frequencies in excess of 30 kHz, offer high spatial-temporal resolution and useful sensitivity to depths up to 200 m. When vessel mounted, these sounders can perform rapid surveys of zooplankton scattering layers, and are particularly useful for assessing small-scale patchiness and vertical layering. The drawbacks of purely acoustic surveys are an uncertain knowledge of the scatterer type (i.e., species and size), combined with an uncertainty regarding appropriate acoustic scattering models, making it difficult to estimate the absolute abundance using acoustics alone. When combined with net trawls and/or new *in situ* optical tools, however, accurate surveys of zooplankton abundances are feasible (Wiebe *et al.*, 1997; Greene *et al.*, 1998). These *in situ* samples provide the necessary data on the species, sizes, orientation, and abundances at specific depths or locations. The link between acoustic and *in situ* sampling approaches is a reliable acoustic scattering model for a particular zooplankton type. Of course, *in situ* sampling is more difficult, the postanalysis of net samples is time-consuming, and both nets and optical techniques lack the *volumetric coverage rate* of the echo-sounders. Therefore, the thrust of this work is to demonstrate a combined acoustic vs *in situ* measurement approach, including a validation of generic scattering models for two zooplankton types.

Quantitatively, the acoustic backscatter strength is related to the zooplankter type, orientation, and their abundance within the insonified volume. It is well known that a *single*-frequency echo-sounder cannot generally distinguish between mixtures of different zooplankton sizes or species (e.g., Greene *et al.*, 1989; Holliday and Pieper, 1993). Similarly, it has been found that single-frequency systems have difficulties distinguishing between zooplankton and colocated turbulent microstructure scattering (Stanton *et al.*, 1994; Trevorrow, 1998). Thus, single-frequency systems are only useful in nonturbulent regions dominated by a single species, or in situations where different species can be distinguished on the basis of behavior (e.g., depth). Such situations were found in this study. A more general approach is the simultaneous use of multiple acoustic frequencies, typically 30 to 1000 kHz, as demonstrated in several previous

[a)]Electronic mail: mark.trevorrow@drdc-rddc.gc.ca

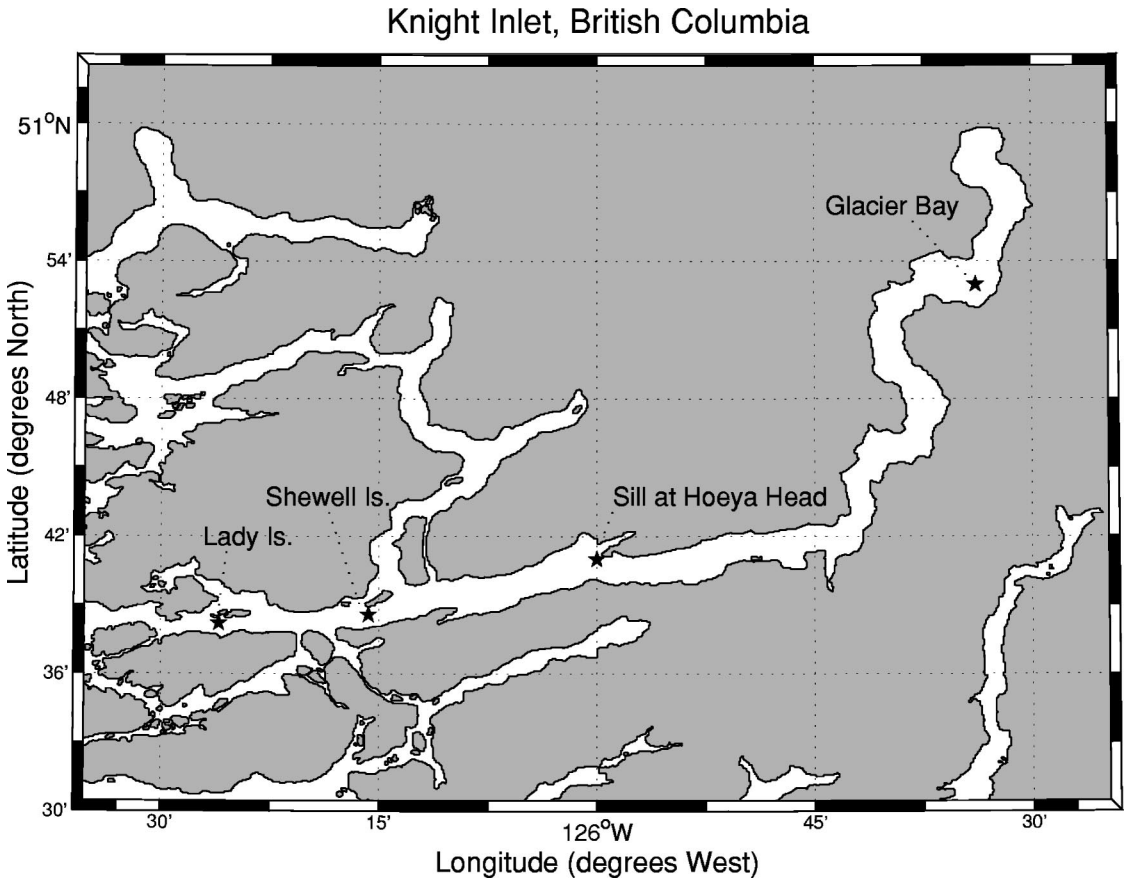


FIG. 1. Map of Knight Inlet showing locations of data collection sites.

studies (e.g., Greenlaw, 1979; Holliday and Pieper, 1980; Greene *et al.*, 1989; Trevorrow and Tanaka, 1995; Wiebe *et al.*, 1997). Because the scattering functions for crustacean zooplankton are strongly size-and frequency dependent, variations in scattering strength with frequency can be used to infer animal sizes. Without *in situ* data on zooplankton species and sizes, however, appropriate scattering models cannot be chosen with certainty.

This study combined a three-frequency, vessel-based echo-sounder system with BIONESS multinet plankton trawls and a newly developed *in situ* optical imaging system known as ZOOVIS. The surveys were performed at various locations in Knight Inlet, British Columbia (B.C.) as part of an investigation on mechanism for zooplankton aggregation at coastal sills. Two separate 15-day sea trials were conducted in November 2001 and 2002. Knight Inlet has been the site of several recent oceanographic experiments (e.g., Farmer and Armi, 1999a, b; Klymak and Gregg, 2001, 2003, 2004) focused on internal hydraulics, and turns out to be a unique natural laboratory for study of zooplankton abundance and aggregation. This work will demonstrate several examples where the combination of acoustic versus *in situ* samples can be used to verify generic zooplankton scattering models.

II. EXPERIMENTAL DESCRIPTION

A. Overview of the physical and ecological setting

Knight Inlet is a fjord located on the south-central coast of British Columbia, fed by a glacier at its head some 90 km inland, and open to the coastal Pacific through Queen Charlotte Sound at its mouth. Knight Inlet is relatively calm, with comparatively light winds and largely sheltered from oceanic wave action, making it ideal for survey operations. The lower part of Knight Inlet trends approximately WSW and has a roughly uniform width (see Fig. 1), whereas the upper half of the Inlet trends more north–south. The inlet is quite steep-sided, with side-wall slopes near 45° and mountain peaks on each side reaching over 1000 m above sea level. In the middle of the lower section there is an underwater sill running approximately north–south, roughly perpendicular to the east–west tidal flow, with typical crest depth near 65 m. This sill region has been the focus of extensive internal hydraulics investigations (Farmer and Armi, 1999a, b; Klymak ans Gregg, 2001, 2003, 2004; Ross and Lueck, 2003). To the west of the sill the seabed falls away to a flat bottom roughly 150 m deep, while on the eastern side and in the upper inlet there is a deeper basin with depths more than 500 m.

The area has dominantly semidiurnal tides, with a strong fortnightly modulation and tidal heights ranging from 1.0 to 4.5 m. The flood tide runs eastward and northward. Typical peak tidal currents over the sill are near $1.5\,\mathrm{m\,s^{-1}}$. During the late autumn Knight Inlet features a strong near-surface stratification, with a colder, fresher (7.5 °C, 26-psu) surface layer from 6–10-m deep with relatively well-mixed conditions (8.0 °C, 32 psu) in the deeper waters. Vertical gradients of temperature, salinity, and density within the 2–5-m thick pycnocline were typically strong (maxima approximately

$0.06\,^{\circ}\mathrm{C\,m^{-1}}$, $1.6\,\mathrm{psu\,m^{-1}}$, and $1.8\,\mathrm{kg\,m^{-4}}$). The deep waters (>50 m) in the eastern basin and upper inlet tend to be slightly fresher than on the west side of the sill.

Ecologically, Knight Inlet is similar to other B.C. coastal inlets. In these inlets zooplankton scattering layers are commonly observed at 60–90-m depth in the daytime, exhibiting clear nocturnal migration to the near surface. These layers are generally composed of larger crustacean zooplankton such as euphausiids (dominated by *E. pacifica*), amphipods (both hyperiid and gammarid), copepods, and various shrimp (Mackie and Mills, 1983). Other species of pteropods, chaetognaths, ctenophores, and cnidaria are sometimes present, and known to migrate diurnally from the surface through depths of 250 m. However, in the context of interpreting echo-sounding data these other species can be largely ignored due to the combination of their low abundance and small target strength (these are soft-bodied animals). The exception to this is the hard-shelled planktonic pteropod *Limacina helicina*, which for typical animal size near 3 mm have a target strength at 200 kHz near -76 dB (Stanton *et al.*, 1998), similar to that of adult euphausiids. A few *L. helicina* were collected in the BIONESS trawls, with abundance ranging from zero to a maximum of $6.8\,\mathrm{per\,m^3}$. All examples discussed in this work had negligible abundances of *L. helicina*.

The most numerically abundant zooplankton type in the inlet are copepods, with adults reaching up to 4 mm in length. *Neocalanus plumchrus* appears to be the most common species in B.C. coastal waters (Harrison *et al.*, 1983). *Neocalanus spp.* generally do not migrate on a daily basis, and in the midsummer the adults dive to deep waters to overwinter (diapause) and thus were not present in these November surveys. Other smaller copepod species (*Calanus marshallae, C. pacificus, Pseudocalanus spp.*, and *Metridia pacifica*) are common in B.C. coastal waters, including Knight Inlet, and these generally exhibit similar seasonal migration. Owing to the combination of the small size of the adult copepods relative to the euphausiids and amphipods, and their relatively low abundances in the scattering layers ($\mathrm{typically} < 100\,\mathrm{per\,m^3}$), scattering from copepod species is ignored in this work.

In contrast to the crustacean scattering dominant in the lower inlet, acoustic scattering signatures in the upper inlet appeared to be dominated by the physonect siphonophores. This type of siphonophore is a relatively strong acoustic scatterer due to the presence of a small gas-filled bubble (the pneumatophore) at the top of the colony. Unfortunately, trawl nets usually destroy these delicate siphonophore colonies, such that abundances can only be inferred from counting remaining component parts (e.g., nectophores). Using visual observations from a submersible, Mackie (1985) reported significant abundances of *Nanomia cara* in Knight Inlet and other B.C. coastal inlets.

B. Instrumentation and methods

The primary acoustic tool used during the Nov. 2002 surveys was a three-frequency (38, 120, and 200 kHz) echo-sounder system mounted on the *CCGS* VECTOR. This system was utilized to both map out the spatial distributions of meso-zooplankton, and for detailed comparison with the BIONESS and ZOOVIS. The 120-kHz sounder proved to be the most useful in this study, as it was sensitive to both fish and zooplankton aggregations and the backscatter from turbulent microstructure that is a feature of the internal flow regime near the sill.

During the Nov. 2002 surveys the single-beam three echo-sounder transducers were mounted in a single fairing at a depth of 3.5 m approximately 25 m forward of the stern. The transducers were oriented vertically downward. The beamwidths of the transducers were approximately matched, being 9.5°, 9.0°, and 7.0° (total width to -3 dB) for the 38-, 120-, and 200-kHz transducers respectively, so that the overall insonified volumes were similar. Transducer beam patterns were provided by the manufacturer. The echo-sounder transmitter–receiver electronics were custom built at the Institute of Ocean Sciences (Sidney, B.C.), with echo-sounder control, data acquisition, and display performed by an accompanying PC. During all survey runs the echo-sounder was operated at a 1.0-Hz ping rate, with all channels simultaneously transmitting a 0.5-ms (37-cm acoustic resolution) gated-cw pulse. The acoustic source levels for all three transducers were near 221 dB ($re:\mu$Pa at 1 m). At this source level acoustic cavitation effects were not considered to be a problem due to the relatively high acoustic frequencies, short pulse length, and very low duty cycle (see Medwin and Clay, 1998). Echo intensity from each channel was sampled at 12 500 samples per second (6-cm sampling resolution) with 16-bit (93-dB) dynamic range. The nominal maximum recording depth was 200 m. In addition to the acoustic data, time and DGPS position were recorded for each ping. During echo-sounder surveys the ship maintained a speed of roughly 6 knots ($3\,\mathrm{m\,s^{-1}}$) relative to the ground, except during BIONESS trawls where the speed was reduced to roughly 3 knots and during ZOOVIS deployments where the speed was 1 to 2 knots.

Sea-truth samples of zooplankton abundance, size, and species composition were obtained by towing a BIONESS (Bedford Institute of Oceanography Net and Environmental Sampling System) instrumented multiple net sampler (Sameoto *et al.*, 1980) at selected locations. The BIONESS carried nine nets that were opened in sequence. Each net had a mouth area of $0.25\,\mathrm{m^2}$ and a mesh size of 0.23 mm. Depth and cumulative volume filtered were continuously monitored with a pressure sensor and flow-meter interfaced to a Seabird Electronics CTD mounted on the BIONESS. The BIONESS was either towed horizontally or obliquely at a forward speed of about $1.5\,\mathrm{m\,s^{-1}}$. The nets were opened and closed in sequence, either to divide the water column into a stacked series of depth strata, or to obtain a horizontal sequence of samples from one depth stratum (e.g., tracking an euphausiid scattering layer). The echo-sounders were used during the BIONESS tows to guide the nets to particular zooplankton scattering layers, and to prevent impact with the seabed. The entire sample from each BIONESS tow stratum was preserved in 10% formalin and returned to the laboratory for identification and enumeration. The entire sample was examined for abundance and body size of large and/or rare taxa

(e.g., euphausiids, amphipods, gelatinous zooplankton). The sample was then quantitatively subsampled using a Folsom splitter for counts of small abundant taxa (e.g., copepods). Total counting effort was sufficient to give an expected subsampling error of roughly 5%–10% for the dominant species.

The Zooplankton Visualization and Imaging System (ZOOVIS) is a high-resolution digital camera system designed to collect quantitative data on the numbers, sizes, and identities of meso-and macro-zooplankton at depths up to 250 m. A detailed description of this system can be found in Benfield *et al.* (2003a). ZOOVIS was capable of resolving anatomical details of organisms larger than 1–2 mm in size within a sampling volume approximately 40 cm away from the camera. Sampling volumes ranged from 0.36 to 3.4 liters depending on camera settings. Images were acquired at 4–8-s intervals while the instrument was profiled up and down at $50\,\mathrm{cm\,s^{-1}}$ with the vessel steaming at 1–2 knots. Similarly to BIONESS, real-time echo-sounder images were used to guide ZOOVIS to particular scattering layers. Due to shipboard logistical constraints, however, ZOOVIS and BIONESS could not be deployed simultaneously. Images were analyzed through a combination of visual inspection and postprocessed with an image processor that located and measured targets. Euphausiid sizes were estimated by measuring their widths and converting to length using an allometric relationship determined from preserved *Euphausia pacifica* collected on 19 Nov. 2002. Euphausiid abundances were estimated by summing all detected euphausiids and dividing by the corresponding imaging volume.

C. Acoustic postprocessing

For quantitative purposes, the raw echo-sounder amplitude vs range data, $A(r)$, must be converted to volumetric backscatter strength, S_v, defined as the decibel equivalent of the backscatter cross section per unit volume (units of dB $re:\mathrm{m}^{-1}$). This conversion is accomplished through the standard sonar equation (e.g., Medwin and Clay, 1998), i.e.,

$$S_v(r) = 20\log_{10}[A(r)] + K + 40\log_{10}[r] + 2\cdot\alpha\cdot r - 10\log_{10}[U(r)] - \mathrm{TVG}(r), \quad (1)$$

where K is the calibration coefficient, r is downward range, α is the acoustic absorption, $U(r)$ is the insonified volume, and TVG is the preamplifier time-varying gain. The calibration coefficient K includes transmit power level, transducer sensitivity, fixed amplifier gain, and A/D conversion factors. The echo-sounder TVG approximated a $20\cdot\log_{10}[r]$ function within the 60 ms after transmission and was held constant at greater depths. The TVG was measured for each channel by averaging echo-data with the transmitter turned off, then normalizing such that TVG at $r=1$ m was zero. The TVG was not changed during these field trials. The acoustic absorption is frequency- and environment dependent, and for the typical (depth-averaged) water conditions in Knight Inlet has a values of 0.0093, 0.033, and $0.047\,\mathrm{dB\,m^{-1}}$ at 38, 120, and 200 kHz, respectively (Francois and Garrison, 1982). In terms of the *equivalent solid angle* of the transducer, φ, and the pulse duration, $\tau(=500\ \mu\mathrm{s})$, the insonified volume is given by

$$U(r) = \frac{1}{3}\varphi\left[\left(r+\frac{1}{4}c\tau\right)^3 - \left(r-\frac{1}{4}c\tau\right)^3\right]. \quad (2)$$

For the sounders used in Knight Inlet the insonified volume increased from roughly $0.5\ \mathrm{m}^3$ at 10-m range to $30–50\ \mathrm{m}^3$ at 80 m (the nominal zooplankton layer depth). Within the zooplankton layers these sampling volumes are comparable to that sampled with the BIONESS, but much larger than sampled per image by ZOOVIS.

The three echo-sounders were acoustically calibrated using the backscatter from precisely machined tungsten-carbide spheres (38-, 40-, and 42.9-mm diameter) suspended along the center line of the acoustic beams (following methods outlined in Vagle *et al.*, 1996). Multiple calibration trials were conducted while the ship was anchored at night. The calibrations were estimated to be accurate to ± 0.8 dB, or a relative error of 20% in cross section. The greatest source of error in this was uncertainty in alignment of the target spheres on the transducer axes. The calibration data were carefully selected by moving the sphere through the transducer beams until maxima were found; however, some instabilities in echo amplitude attributable to target sphere movement were observed. Clearly, more accurate calibrations can be obtained using split-beam sounders which can measure the target location relative to the beam.

The electronic noise levels of the echo-sounder electronics created some limitations for zooplankton sampling, especially for the 200-kHz channel at greater depths. Because of the range variations in Eq. (1), these noise levels were also range dependent. Specific systemic noise curves were measured for each channel by recording and averaging echo data with the transmitters turned off. These noise curves are included in all following plots of S_v vs depth to show the measurement signal-to-noise ratio. In all quantitative S_v calculations, usually involving averaging in depth and ping, these noise levels were used to correct for noise contamination, as described in Korneliussen (2000) and Watkins and Brierley (1996). The correction was applied by simply subtracting the equivalent volume backscatter cross section of the noise from the volume scattering cross-section data on a ping-by-ping basis. For moderate to high signal-to-noise data this correction to the averaged S_v is small (<0.5 dB); however, corrections up to 2.5 dB were found in 200-kHz data near the bottom of some scattering layers. An additional signal conditioning process was applied to remove contamination from occasional, large targets (identity unknown) embedded in the zooplankton layers. During ping averaging through zooplankton scattering layers, echoes exceeding an upper S_v threshold in each channel were removed. These upper thresholds were set 10 dB above the maximum layer-averaged S_v in each channel, typically -80 to -70 dB depending on the situation. This contamination was most apparent in crustacean scattering in the 38-kHz channel, where the expected S_v were <-90 dB.

III. CRUSTACEAN AND SIPHONOPHORE SCATTERING MODELS

To proceed from acoustic backscatter strength to *in situ* abundance estimates, several basic facts need to be estab-

lished. The local species and size composition needs to be determined (usually with a net sample), and this species and size composition is assumed to be constant throughout a local survey region. For the purposes of verifying scattering models it is preferable that the acoustic scattering be dominated by a particular species or scatterer type. If the scattering is due to multispecies ensembles, then the composition ratios between major constituents should be reasonably well known. In the case of Knight Inlet, the BIONESS trawls showed that the acoustic scattering in the lower inlet were generally dominated by euphausiids with length near 16 mm, as they are larger and generally in greater abundance than other species. Other zooplankters, such as amphipods, adult copepods, pteropods, and siphonophores, were sometimes found in moderate quantities, and at low euphausiid abundances they sometimes gave significant contributions to the acoustic backscatter. In the upper Inlet (Glacier Bay), the backscattering was dominated by physonect siphonophores, and the net trawls found negligible quantities of euphausiids and amphipods. Three distinct cases will be examined separately in Sec. IV.

Following from the assumption of a particular scatterer type, an accurate model for the acoustic scattering strength as a function of acoustic frequency and zooplankter size must be found. For euphausiids and amphipods, a number of simple models have been proposed (e.g., Johnson, 1977; Stanton, 1989), but the most appropriate in this situation is a size- and orientation-averaged fluid cylinder model recommended by Stanton and Chu (2000), based on earlier models by Stanton *et al.* (1993). This model is appropriate due to the moderately large insonified volumes of these echo-sounders (typically near $30-50\ \mathrm{m}^3$ at the euphausiid layer depth of 60–80 m), such that the echo has contributions from potentially hundreds of animals. Furthermore, the detailed acoustic scattering models generally show a strong sensitivity to zooplankter orientation (i.e., swimming angle with respect to horizontal, i.e., angle = zero implies dorsal incidence), and generally it should be expected that a natural population would have some distribution of this orientation angle about a mean value.

The specific model used here assumes the euphausiids or amphipods to be bent fluid cylinders with radius of curvature $3\times$ body length, a specific length-to-radius ratio, a Gaussian-distributed length (i.e., characterized by a mean and standard deviation), and a Gaussian-distributed orientation angle, again quantified by a mean and standard deviation. The sound-speed and density contrasts in this fluid model were assumed to be 1.03 and 1.04, respectively, following recommendations in Foote *et al.* (1990). Given all these assumptions in combination with net trawl data, the averaged backscatter cross section (σ_b, units of m^2) per animal can be calculated using the modeling equations outlined in Stanton and Chu (2000). One limitation in this formulation is the requirement that the orientation distribution include lateral (dorsal) incidence. The data on euphausiid and amphipod length and aspect ratio come from BIONESS trawl data collected near the sill and at several sites in the western inlet. Also, Stanton *et al.* (1993) showed that this averaged model is insensitive to the amount of curvature, so only a nominal value is assumed. We shall follow the notation suggested by Stanton and Chu where N (mean, std. dev.) specifies the distribution of zooplankter orientation angle (assumed Gaussian, zero = horizontal). A limiting case of orientation angle, denoted *uniformly distributed*, occurs where there is no preferred orientation and averaging occurs over all possible orientations from 0° to 360°. Based on results from Chu *et al.* (1993), the swimming angle distribution was kept as a free parameter to allow some inference of orientation from a best fit to abundances from the BIONESS trawls. Ideally, a full verification of the model could be accomplished through inclusion of *in situ* animal orientations from ZOOVIS. Unfortunately, in most cases examined in this work insufficient numbers of animals were sampled by ZOOVIS to establish reliable swimming angle distributions.

The backscatter from physonect siphonophores is dramatically different from the crustacean model discussed above. The presence of the gas-filled pneumatophore yields a much higher scattering strength than equivalent fluid models, and the radial oscillations of the approximately spherical bubble give rise to a resonant peak. One feature of siphonophores is their apparent ability to maintain a constant pneumatophore volume over a large range of depths, thereby maintaining a constant buoyancy (Benfield *et al.*, 2003b). As a first approximation, the backscatter from the pneumatophore can be modeled as a bubble (Stanton *et al.*, 1998; Warren *et al.*, 2001; Benfield *et al.* 2003b), for which there are well-established models. The backscatter from other fluid-like tissue in the siphonophore colony is significantly weaker than the bubble scatter and can be ignored. This work will follow the textbook formulation for bubble scattering as outlined in Medwin and Clay (1998), whereby the backscatter cross section for a bubble of radius, a, is given by

$$\sigma_b(f,z)=\frac{a^2}{[(f_R/f)^2-1]^2+\delta^2}, \tag{3}$$

where δ is the bubble total damping coefficient, and the bubble resonance frequency is given by

$$f_R=(2\pi a)^{-1}\sqrt{3\gamma P_0/\rho_w}, \tag{4}$$

where P_0 is the static pressure (Pa) on the bubble ($1.013\times10^5+\rho_w\cdot g\cdot z$), γ is the ratio of specific heats at constant pressure and volume of the enclosed gas ($=1.4$ for air), and ρ_w is the density ($\mathrm{kg\,m}^{-3}$) of the surrounding water. For example, a 1-mm-diameter bubble at 50-m depth has a resonant frequency of 15.8 kHz. Note that the total damping coefficient is the sum of three separate terms, i.e., damping due to reradiation, fluid viscosity, and surface tension (see Medwin and Clay, 1998 for specific evaluation of these terms for natural bubbles). It is speculated that the tissue surrounding the bubble in the pneumatophore will exhibit stronger viscous and surface tension damping than for an ordinary bubble in seawater, and the *in situ* scattering data examined here allow some exploration of this idea.

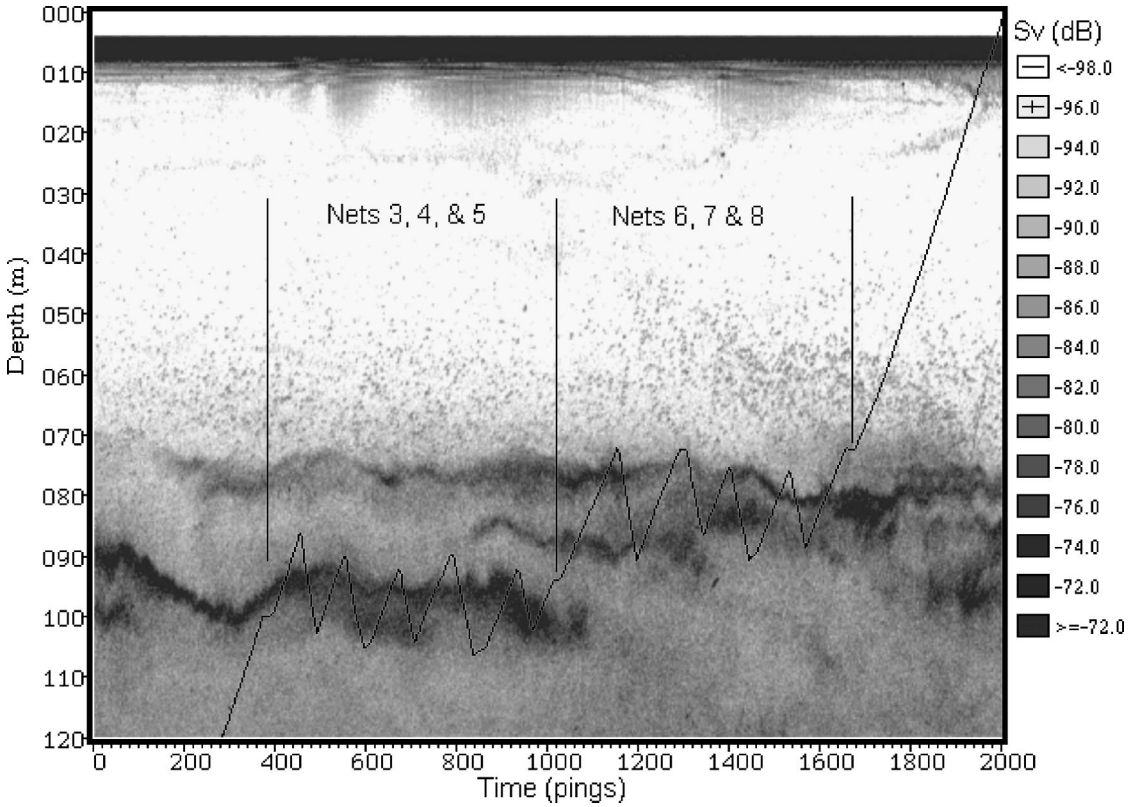


FIG. 2. Volumetric scatter strength vs depth and time at 120 kHz for 33.3 min starting 0833h, 23 Nov. 2002 near Shewell Island in Knight Inlet. Black line shows the trajectory of the BIONESS multinet sampler, with labels indicating zones of net opening.

IV. MULTIFREQUENCY ACOUSTIC RESULTS AND COMPARISONS WITH *IN SITU* DATA

For the purpose of examining the validity of the combined echo-sounder and *in situ* sampling approach, in essence verifying the use of the acoustic scattering models described above, six specific examples from the November 2002 surveys will be examined in detail. These examples are chosen for both the coincidence of good quality acoustic and BIONESS or ZOOVIS data, and the dominance of a particular zooplankton type or species.

A. Euphausiid aggregations in the lower inlet

Scattering layers dominated by euphausiids were commonly observed in the lower inlet and near the sill at Hoeya Head. A specific deep-water comparison between the echosounder and BIONESS results taken at 0830h 23 Nov. 2002 near Shewell Island will be examined here in detail. This example was chosen because it was located well away from the complicated, turbulent flows in the sill region, and because the net trawls indicated a moderate euphausiid abundance with only minor amounts of other species. ZOOVIS was not deployed at this site because earlier experiences had found very few euphausiids were imaged in these relatively low abundance layers.

Figure 2 presents a volumetric backscatter strength vs depth and time echogram, overlain with the trajectory of the BIONESS sampler (appropriately time shifted to compensate for the cable lay-back). The echogram shows the typical daytime scattering layer below roughly 70-m depth. The BIONESS acquired three nets in the lower sublayer at 85–105-m depth and three more nets within the upper sublayer at 72–90 m. Each net sampled between 50 and 80 m^3 within the zooplankton layer. Individual net results within each sublayer exhibited only minor differences, so euphausiid abundances vs size (1-mm size classes) were averaged over each set of three, with result presented in Fig. 3. The measured size distributions are clearly in agreement with the Gaussian-distributed assumption required for the averaged fluid cylinder scattering model described above. For modeling purposes a mean length of 15.9 mm and 1.4-mm standard deviation will be used. Also produced from BIONESS were

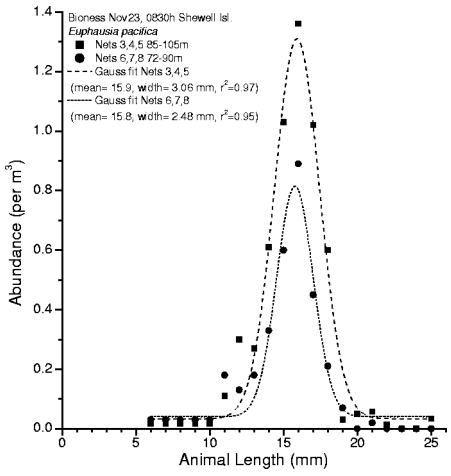


FIG. 3. Size-abundance distribution of *Euphausia pacifica* from BIONESS trawl at 0830h 23 Nov. 2002 near Shewell Island. Best-fit Gaussian distributions plotted as dashed and dotted lines for the two sublayers.

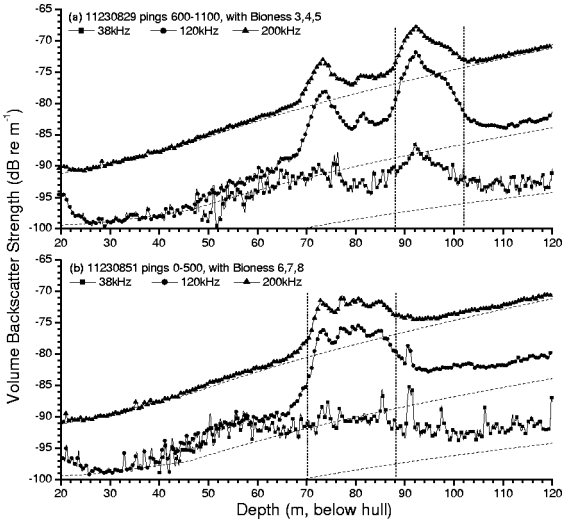


FIG. 4. S_v vs depth profiles at 38, 120, and 200 kHz, averaged over 500 pings (8.3 min) starting at (a) 0839h and (b) 0851h on 23 Nov. 2002 near Shewell Island. Depth-varying dashed lines are corresponding systemic noise levels for the three channels (S_v data are not corrected for noise). Vertical dashed lines show depth intervals of the corresponding BIONESS net trawls, found to be dominated by *E. pacifica* (as shown in Fig. 3).

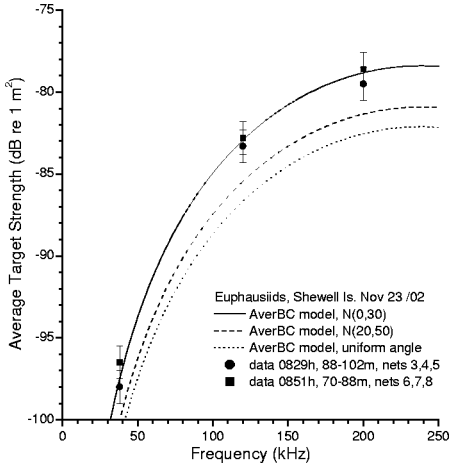


FIG. 5. Comparison between measured and predicted averaged target strength vs frequency for *E. pacifica* near Shewell Island, 0830h 23 Nov. 2002. Experimental TS derived from averaged S_v at 88–104 m and 72–90 m-depth divided by the mean abundance estimated from the BIONESS trawls. Error bars on TS are ± 0.8 dB. Predicted TS calculated from population- and orientation-averaged model using euphausiid length and standard deviation of 15.9 and 1.4 mm, respectively, with three different orientation angle distributions.

averaged total abundances of 5.6 and 3.2 euphausiids per m^3 for the lower and upper sublayers, respectively. This averaged abundance is considered low, as abundances in excess of 500 per m^3 were observed in other parts of the inlet. Finally, an examination of euphausiids captured at several sites in the lower inlet determined that the mean aspect ratio (length to width) was 7.6 ± 0.7. This implies a length to radius ratio of 15.2, similar to that reported for Antarctic krill (*E. superba*) by Foote *et al.*(1990) and Chu *et al.* (1993).

Averaged profiles of volume backscatter strength at the three frequencies for the two zooplankton sublayers are presented in Fig. 4. These volumetric backscatter cross-section profiles were ping averaged over the same interval as the BIONESS net samples shown in Fig. 2, and are uncorrected for systemic noise. These profiles exhibit the distinctive increasing scattering strength with frequency signature of euphausiid scattering layers. Specifically, the ratio between the scattering strength at 120 kHz to that at 38 kHz was approximately 14 dB, and between 200 and 120 kHz the ratio was roughly 4 dB. Within the euphausiid scattering layers the signal-to-noise ratios were generally good (>6 to 10 dB); however, the 200-kHz channel was increasingly noise dominated near the upper and lower boundaries of the euphausiid layers. In general, the 200-kHz channel was not useful for zooplankton sampling at depths >100 m. Note in Fig. 4(b) that the 38-kHz profile does not exhibit the same characteristic increase in scattering strength above the background as observed in the 120- and 200-kHz profiles. This is partly due to the inclusion in the averaged scattering level of echoes from a few larger, isolated targets in combination with low euphausiid signal level. The use of an upper threshold in the depth- and ping-averaging process removes much of this contamination. Quantitatively, within the lower sublayer [88–102 m in Fig. 4(a)], the noise-corrected depth-averaged volume scattering strengths were -90.5, -75.8, and -72.0 dB (*re*: 1 m^{-1}) at 38, 120, and 200 kHz, respectively. Similarly, in the upper sublayer [70–88 m in Fig. 4(b)] the noise-corrected depth-averaged S_v were -91.4, -77.7, and -73.5 dB.

By dividing the averaged volumetric scattering cross section by the measured abundance estimates from the BIONESS within each sublayer, estimates of the averaged target strength (TS, in dB *re*: 1 m^2) per animal at each frequency can be made (Fig. 5). Overall, there is good agreement in the TS estimates between the two sublayers at all three frequencies. Furthermore, there is excellent agreement between the TS estimates and the averaged bent fluid cylinder model prediction using a N(0°,30°) orientation distribution, and there is little overlap with either the N(20°,50°) or the uniform angular distribution predictions. These TS and orientation results are similar to those found for other naturally occurring euphausiid swarms, bearing in mind that most other studies were focused on larger euphausiids. For example, Wiebe *et al.* (1990) reported an average TS of -77.5 dB at 420 kHz for live, tethered, 20-mm-long *E. pacifica*. Foote *et al.* (1990) reported average TS for 34-mm-length *E. superba* near Antarctica in the range -89 to -85 dB at 38 kHz and -79 to -76 dB at 120 kHz. Foote's estimates are 8–12 dB larger at 38 kHz and 4–7 dB larger at 120 kHz for a mean animal size about twice these Knight Inlet euphausiids, entirely consistent with the present fluid-cylinder model using the larger animal size.

In terms of euphausiid orientation, these results are consistent with some previous studies. Based on *in situ* photographs of euphausiids in Norway, Kristensen and Dahlen (1986) reported a mean angle between $\pm 10°$ (dependent on vertical migration state) with standard deviation between 25° and 45°. They concluded that a $N(0°,30°)$ orientation distribution was appropriate for modeling purposes. However, in an examination of Antarctic krill data, Chu *et al.* (1993) extracted swimming angles with mean near 20° (head up) and standard deviation near 30°. Earlier photographic work

by Kils (1981) with Antarctic krill showed a mean orientation of 45° (head up) with a standard deviation of 30°. In tank experiments with similar-sized *E. pacifica*, Miyashita *et al.* (1996) reported typical swimming angle distributions with mean of 30° (head up) and standard deviation of 20°.

The BIONESS trawl data can also be used to assess the error due to scattering contributions from other species. Aside from euphausiids, the next most abundant zooplankton class was amphipods, with measured length distribution of 6.3 ± 2.4 mm and averaged abundance of 6.6 m^3 in both the lower and upper sublayers. Using a similar size and orientation-averaged scattering model (amphipod parameters are described in the next subsection), the predicted layer-averaged S_v values for amphipods are -100.5, -88.3, and -83.5 dB (*re*: 1 m^{-1}) at 38, 120, and 200 kHz, respectively. These are all more than 10 dB lower than the observed S_v attributed to euphausiids, and thus amphipod scattering can be guardedly ignored.

B. Amphipods in the western inlet

A combined acoustic and BIONESS survey at 1125h on Nov. 19, 2002 near Lady Island, in the western part of the lower inlet, found a situation dominated by amphipod scattering with almost negligible presence (abundance <0.1 per m^3) of euphausiids. During the BIONESS trawl, four individual nets were deployed within main scattering layer at 60–80-m depth, with an averaged abundance of 22.9 per m^3. Each net sampled between 80 and 100 m^3, so this abundance estimate is an average over roughly 330 m^3 of the scattering layer. There were two dominant species of amphipods, namely a hyperiid *Themisto pacifica* and a gammarid *Cyphocaris challengeri*. The combined size distribution of the hyperiid and gammarid amphipods was found to be Gaussian distributed with mean of 6.3 mm and standard deviation of 2.4 mm. The mean aspect (length-to-width) ratio measured from a number of amphipod samples was 5.2. ZOOVIS was not deployed at this site.

Ping-averaged volume backscatter strength profiles through this amphipod layer are presented in Fig. 6. Similar to euphausiid layers discusssed earlier, the three frequencies all exhibited a strong S_v signal within the 58–76-m amphipod layer and an increasing S_v with frequency. The noise-corrected depth- and ping-averaged S_v were -93.7, -83.2, and -79.0 dB (*re*: 1 m^{-1}) at 38, 120, and 200 kHz, respectively. As in the euphausiid case (Fig. 4), the 200-kHz profile had reasonable signal-to-noise properties within the amphipod scattering layers, but was clearly noise limited outside of the layer. Also, the ping-averaged 38-kHz S_v profile exhibited a different scattering texture, with rapid fluctuations in depth, both within and outside of the amphipod layer. This is in contrast to the more continuous, smooth scattering profiles of the 120- and 200-kHz channels. This contamination is due to the inclusion of echoes from a few isolated targets (identity unknown) in the averaged profile, made apparent by the small 38-kHz signature from amphipods. Again, the use of an upper threshold in the depth- and ping-averaging process removes much of this contamination.

Estimates of the averaged TS for these amphipods can be made by dividing the time- and depth-averaged volumetric backscatter cross section at each frequency by the measured abundance estimate from the BIONESS. Figure 7 compares the measured and modeled TS for this amphipod layer. In this case there is agreement with the model using a N(0°,60°) distribution, although the uniformly distributed case cannot be rejected. The 38-kHz estimate appears too high by roughly 2 dB, presumably due to residual contamination from these isolated scatterers mentioned above. Finally, we can use the upper bound on euphausiid abundance (0.1 per m^3) taken from the BIONESS trawl data to assess the potential interference with these amphipod S_v measurements. The fluid-cylinder model for euphausiids (discussed

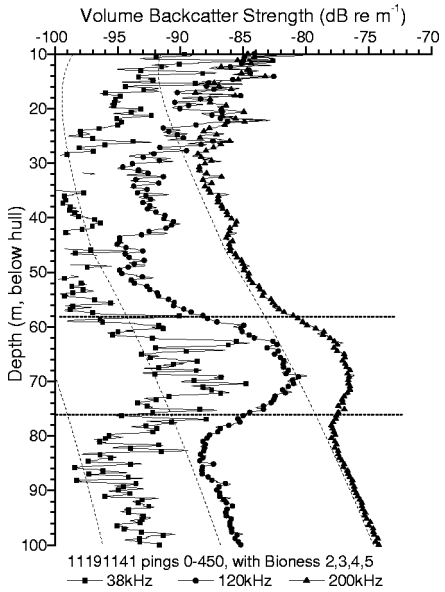


FIG. 6. S_v vs depth profiles at 38, 120, and 200 kHz, averaged over 450 pings (7.5 min) starting 1141h, 19 Nov. 2002 near Lady Island in Knight Inlet. Depth-varying dashed lines are corresponding systemic noise levels for the three channels (S_v data are not corrected for noise). Horizontal dashed lines show depth interval of the corresponding BIONESS net trawls, found to be dominated by hyperiid and gammarid amphipods.

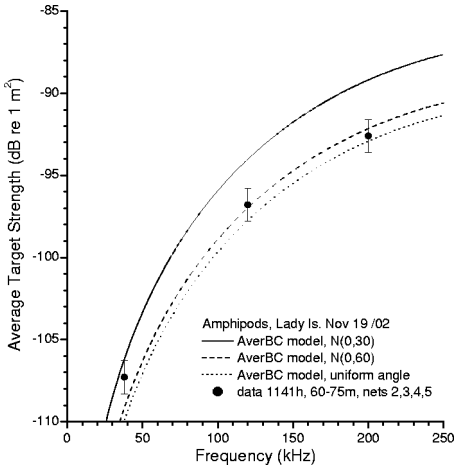


FIG. 7. Comparison between measured and predicted averaged target strength vs frequency for amphipods near Lady Island, 1141h 19 Nov. 2002. Experimental TS derived from averaged S_v at 60–78-m depth divided by the mean abundance estimated from the BIONESS trawl. Error bars on TS are ± 0.8 dB. Predicted TS calculated from population- and orientation-averaged model using amphipod length and standard deviation of 6.3 and 2.4 mm, respectively, with three different orientation angle distributions.

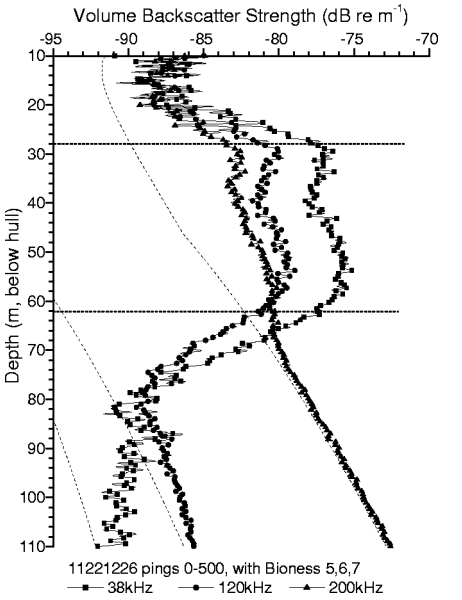


FIG. 8. S_v vs depth profiles at 38, 120, and 200 kHz, averaged over 500 pings (8.3 min) starting 1226h, 22 Nov. 2002 in Glacier Bay, Knight Inlet. Depth-varying dashed lines are corresponding systemic noise levels for the three channels (S_v data are not corrected for noise). Horizontal dashed lines show depth interval of the corresponding BIONESS net trawls, which showed presence of siphonophores and lack of euphausiids or amphipods.

above) predicts averaged S_v in this layer of -107.8, -93.1, and -88.9 dB (*re*: 1 m^{-1}) at 38, 120, and 200 kHz, respectively. These are all more than 11 dB below the observed S_v data for the amphipod scattering layer.

C. Siphonophores in Glacier Bay

Acoustic scattering signatures in the upper inlet were found to be drastically different from those found in the lower inlet. In this case the scattering strength vs frequency trend was reversed, with the 38 kHz being the strongest and the 200 kHz the weakest, as demonstrated in Fig. 8 with an example from midday on 22 Nov. 2002. Clearly, all three echo-sounders were able to resolve the scattering layer with moderate to good signal-to-noise ratios. Focusing on a region near the upper boundary of this layer (28–32 m), the noise-corrected depth- and ping-averaged S_v were -77.1, -80.5, and -83.8 dB (*re*: 1 m^{-1}) at 38, 120, and 200 kHz, respectively. A simultaneous BIONESS trawl from 30–65-m depth detected the presence of siphonophore (identified as *Nanomia bijuga*) nectophores, pneumatophores, and siphonulae. Specifically, within this 30–65-m layer the averaged abundances of nectophores and pneumatophores from BIONESS were 1.29 and 0.13 per m^3, respectively. Since it is reasonable to expect roughly 10 nectophores per *Nanomia* colony but only one pneumatophore, these two estimates are consistent. The trawls also found a distinct absence of euphausiids and amphipods in this layer (averaged euphausiid and amphipod abundances were <0.15 and 0.4 per m^3, respectively). Using the euphausiid and amphipod TS outlined in previous sections, these measured abundances predict scattering strengths that are much smaller than observed, particularly for the 38-kHz channel. ZOOVIS was deployed at this site prior to the BIONESS cast, but did not return any siphonophore images due to the low abundances.

Because of the relative fragility of the siphonophore colonies and their relatively low abundance, standard net trawls (such as BIONESS) are considered somewhat unreliable for measuring siphonophore abundance. Fortunately, an alternate estimate of the average target strength can be derived from an examination of the scattering layer statistics, using a technique outlined in Stanton (1985) and used by Trevorrow and Tanaka (1997) to estimate *in situ* amphipod TS. The basis of this technique, dubbed *critical density analysis*, is that probability density functions of scattering amplitude show different shapes between the two extremes of overlapping and nonoverlapping scatterer echoes. The transition between these two extremes, which is generally rather sharply defined in depth and easy to identify from PDF shape, corresponds to a depth where *on average* there is one scatterer per insonified volume (i.e., a *critical* density). With knowledge of the averaged S_v at this depth of critical density, the average backscatter cross section (i.e., TS) can be extracted.

In the specific example examined here, critical density points for the 38- and 120-kHz channels were found at 29.4-m depth. The corresponding critical densities were 0.162 m^{-3} for both channels, with noise-corrected, ping-averaged S_v of -76.2 and -80.3 dB (*re*: 1 m^{-1}), at 38 and 120 kHz, respectively. Combining the critical density and average scattering strength produced averaged TS estimates of -68.3 and -72.4 dB (*re*: 1 m^2) at 38 and 120 kHz, respectively. A TS estimate at 200 kHz of -75.1 dB was generated using the ratio between layer-averaged S_v at 38 and 200 kHz (-6.8 dB). A similar critical density point was found in the 38-kHz channel at the bottom of the layer (75-m depth), yielding at 38 kHz TS of -68.4 dB (*re*: 1 m^2). Using ratios between averaged S_v at the three frequencies near 75-m depth, the estimated TS at 120 and 200 kHz at this depth was -72.3 and -75.1 dB (*re*: 1 m^2), respectively. A similar critical density analysis performed on a separate data set taken earlier (1020h) on the same day in a nearby location produced similar results (e.g., averaged TS$=-69.1$, -73.1, and -75.5 dB at 38, 120, and 200 kHz at 30-m depth).

Using these TS values, the average siphonophore abundance was near 0.14 per m^3, very close to the abundance estimated from the nectophores and pneumatophores captured in the BIONESS trawl. Similar abundances (mean of 0.05 per m^3, maximum up to 1 per m^3) of *Nanomia bijuga* in Monterey Bay were reported by Robison *et al.* (1998) from ROV-based video observations. However, these present estimates are considerably lower than 1–10 siphonophores per m^3 estimated acoustically for *N. cara* in the Gulf of Maine reported by Warren *et al.* (2001), Benfield *et al.* (2003b), and earlier submersible observations in the same Gulf of Maine region by Rogers *et al.* (1978).

A comparison between these TS estimates and siphonophore scattering models is presented in Fig. 9. Because the scattering model for a bubble (our model for a pneumatophore) is depth dependent, the figure presents comparisons at 30 and 60-m depth, near the upper and lower boundaries of

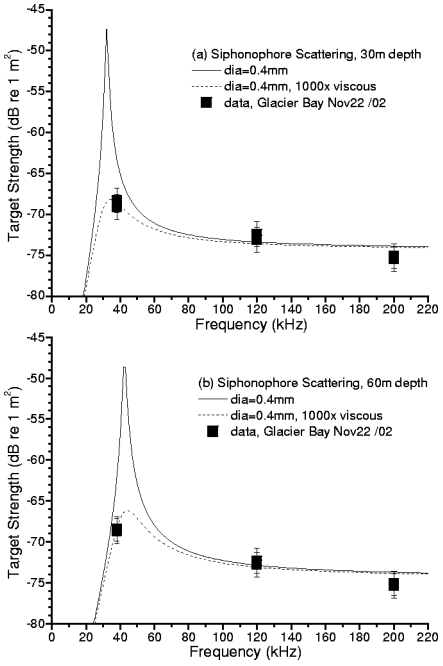


FIG. 9. Comparison between measured and predicted averaged target strength vs frequency for siphonophores in Glacier Bay, Knight Inlet. Experimental TS derived from critical density analysis (as described in the text) from data at 1017h and 1226h, 22 Nov. 2002. Error bars on TS are ± 1.5 dB.

the siphonophore layer. The bubble model clearly exhibits a resonant peak near 40 kHz, and that this resonant frequency increases with depth. The best overall fit at frequencies above the resonance (i.e., 120 and 200 kHz) comes for a 0.4-mm bubble diameter. However, for this size the unmodified bubble model overpredicts the measured TS at 38 kHz at both 30 and 60 m. A better fit to the 38-kHz TS at both depths can be found by arbitrarily increasing the viscous damping component by a factor of 1000. Furthermore, note that at 30-m depth the resonant peak is below the 38-kHz echo-sounder frequency, while at 60-m depth the resonance is above 38 kHz. Assuming the pneumatophore size to be roughly uniform throughout this layer, the unmodified bubble model predicts that a strong resonant scattering region, roughly 20 dB higher, should occur in the 38-kHz channel at some intermediate depth. From Eq. (4), a 0.4-mm-diameter bubble has a predicted 38-kHz resonance at a depth of 45 m. Examination of Fig. 8 indicates that the ratio of S_v between 38 and 120 kHz is approximately constant from 30- to 60-m depth, with no suggestion of a resonant peak near 45 m. These facts lend support to the earlier hypothesis that the membrane surrounding the gas bubble in the pneumatophore adds damping. However, the exact damping contributions are difficult to assess except under carefully controlled laboratory conditions. This 0.4-mm pneumatophore diameter lies in between adult pneumatophore sizes of 0.7–1.0 mm examined in Warren *et al.* (2001), and siphonulae pneumatophore diameters of 0.1-0.4 mm found by Benfield *et al.* (2003). Furthermore, these present measurements are consistent with previous siphonophore TS measurements. For example, Warren *et al.* reported average TS measurements of -62.5 dB at 24 kHz, while Benfield *et al.* predicted average TS of -72 dB at 120 kHz and -77 dB at 200 kHz. Independently, Greene *et al.* (1998) reported a siphonophore TS of -75 dB at 420 kHz based on direct acoustic vs net trawl comparisons in the Gulf of Maine.

D. Echo-sounder vs ZOOVIS comparisons

The occurrence of dense euphausiid scattering layers in the lower inlet also allowed closely coupled comparisons between the echo-sounders and ZOOVIS. For these comparisons, 120-kHz volume scatter strength can be converted to euphausiid abundances using results from Sec. IV A, specifically using a mean TS of -83.1 dB (*re*: 1 m^2). The 120-kHz channel is used as it provided the best overall signal-to-noise performance for the euphausiid layers. However, in contrast to the BIONESS comparisons where time- and depth averages were appropriate, comparisons to the small sample volume of ZOOVIS (of order 1 liter) requires high-resolution acoustic data. For this analysis the echo-sounder data were averaged over two successive pings (2 s) and into 24-cm-depth bins. Figure 10 presents a detailed ZOOVIS vs acoustic comparison for a shoaling near-surface euphausiid layer near Lady Island. Typical acoustically derived abundances in this example are 100 to 200 per m^3, with isolated patches exceeding 500 per m^3. The maximum acoustically predicted abundance is 775 per m^3. Note that the insonified volume of the 120-kHz echo-sounder increases with depth from 0.6–4.3 m^3 within this 8–22-m deep layer, approximately 10^3 to 10^4 times larger than the sample volume of ZOOVIS (0.36 liter in this deployment). The ZOOVIS-derived abundances are based simply on counting the number of recognizable euphausiid targets per image, then dividing by the sample volume. Note that the minimum quantity of 1 euphausiid per image corresponds to an abundance of 2.8 per liter (2800 per m^3), much larger than any observed with the echo-sounder. In this example no more than 1 euphausiid per image was observed, and the majority of images was empty. The overall average ZOOVIS abundance from this deployment was 0.427 per liter. Generally, there is reasonable agreement in the sense that high ZOOVIS abundances correspond to locations where the instrument profiled through denser acoustic regions, and empty ZOOVIS images come from zones with low acoustic abundance.

Overall the ZOOVIS abundances presented in Fig. 10 appear to be about 2 to 4 times higher than the acoustic estimates. There are a number of possible contributing factors to this discrepancy. First, the huge difference in sampling volumes implies different sampling statistics for the two measurements, one being an average over many targets per insonified volume and other having only isolated hits. However, the average ZOOVIS abundance (including zero estimates) is still about a factor of 2 larger than the typical acoustic abundances. Another complication is that of animal swimming-angle orientation. If the euphausiids had a different orientation distribution, say closer to uniformly distributed, then their average TS could be plausibly lower by as much as 3 dB (see Fig. 5). Some evidence for this can be seen from an oblique bongo net trawl over the upper 30 m taken roughly 1 h later at this same Lady Island site. The

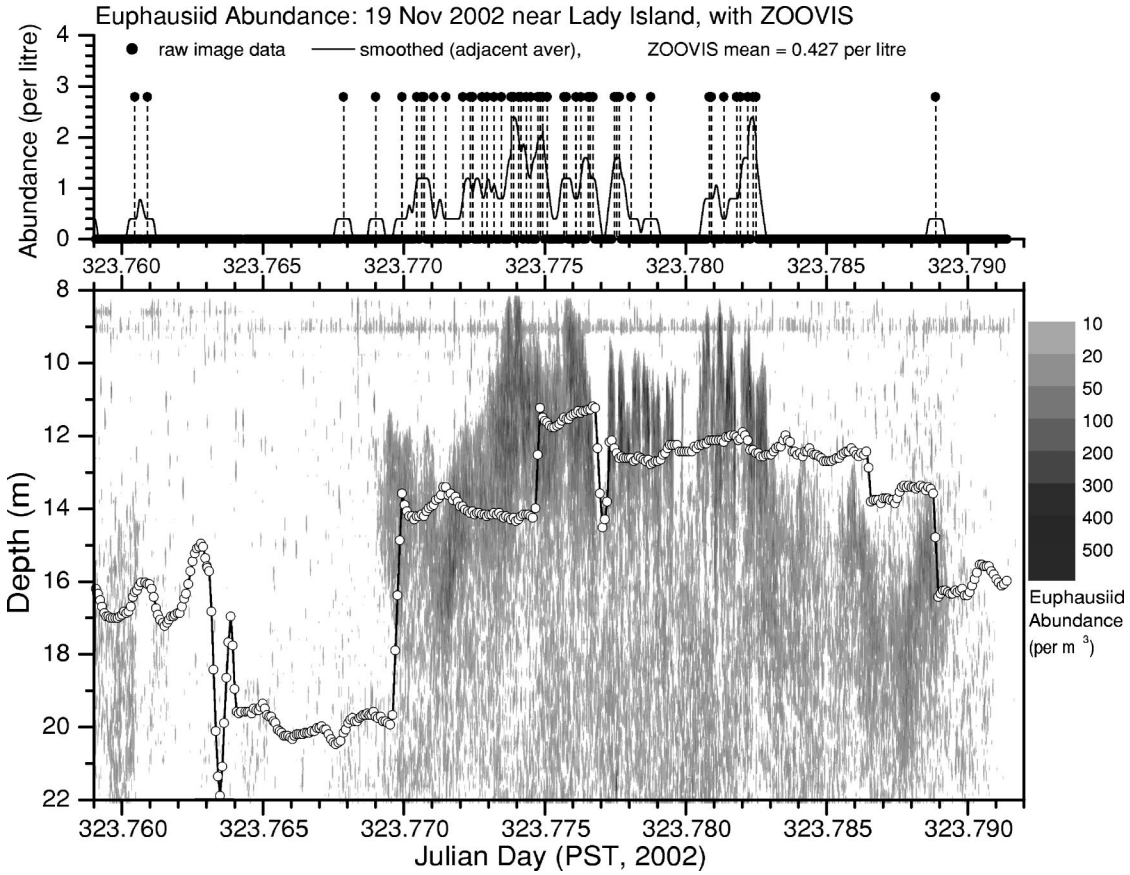


FIG. 10. Comparison of echo-sounder and ZOOVIS abundance time-series, 1812h to 1900h 19 Nov. 2002 near Lady Island. Upper panel shows ZOOVIS abundances (per liter) based on a 0.36-liter sampling volume. Smoothed curve is a 7-point adjacent average. Lower panel shows echo-sounder abundance (per m^3) vs depth and time calculated from 120-kHz echo-sounder assuming 15.9-mm-length euphausiids as described in Sec. IV A. Echo-sounder data has 2-s time and 24-cm depth resolution. ZOOVIS trajectory overlain on echo-sounder contour plot, with circles indicating individual images.

bongo picked up the usual 15.9-mm length euphausiids, with a layer-averaged abundance of 23.5 per m^3. Averaged over the same depth and time interval as the bongo trawl, the acoustically derived abundance estimate (based on the same methods as for the ZOOVIS comparisons) was 16.5 per m^3. This suggests that the average euphausiid TS could be smaller by 1.5 dB, potentially pushing the acoustically derived abundances in Fig. 10 upwards by roughly 50%. A final complication is that this region clearly exhibited high spatial heterogeneity, and it is possible that ZOOVIS was towed through slightly different euphausiid swarms than sampled by the echo-sounder. The difference between the two sampling points was roughly 30-m distance and 20 s in time.

Figure 11 presents another, deeper acoustic vs ZOOVIS comparison near the Hoeya Head sill. This example contains some of the highest euphausiid densities observed during the Nov. 2002 field trail. Acoustically, the typical peak abundance was 200 to 400 per m^3 with several patches in excess of 1000 per m^3. The maximum acoustically derived abundance was 1610 per m^3 for the patch at JD=327.659 and 77-m depth. At these greater depths the acoustically insonified volumes were 32–45 m^3, again roughly 10^4 times larger than the ZOOVIS sample volume (3.4 liters). In this deployment the minimum observable ZOOVIS abundance was 0.29 per liter, and there were several images with multiple euphausiids (up to 8 in one image). The overall average ZOOVIS abundance was 0.36 per liter, in rough agreement with the acoustic estimates, and there is reasonable correlation in the locations of abundance peaks. However, the peak ZOOVIS values (up to 3400 per m^3) were approximately a factor of 2 larger than the acoustics. Due to the increased flow-induced turbulence near the Hoeya Head sill, it is plausible that the euphausiids could have had a different swimming-angle distribution, likely closer to uniformly distributed. This effect could reduce the average TS by roughly 3 dB, effectively doubling the acoustically derived abundance. Additionally, there was a larger spatial separation of the acoustic and ZOOVIS samples created by the greater length of cable required to deploy ZOOVIS to this depth. The estimated separation was roughly 80-m distance and 52 s in time, increasing the potential sampling mismatch.

E. Effects of vessel lights on near-surface euphausiids at night

A dramatic example of euphausiid reaction to light was found during a nighttime survey near the Hoeya Head sill on 21 Nov. 2002. At this time the vessel was moving slowly during a deployment of ZOOVIS. When the aft-deck flood lights were turned on, the euphausiid scattering layer near 20-m depth suddenly dropped in intensity by roughly 10–15 dB, depending on frequency. This change in the scattering layer was found to be repeatable by alternating 2-min periods with the lights on and off. Figure 12 compares scattering

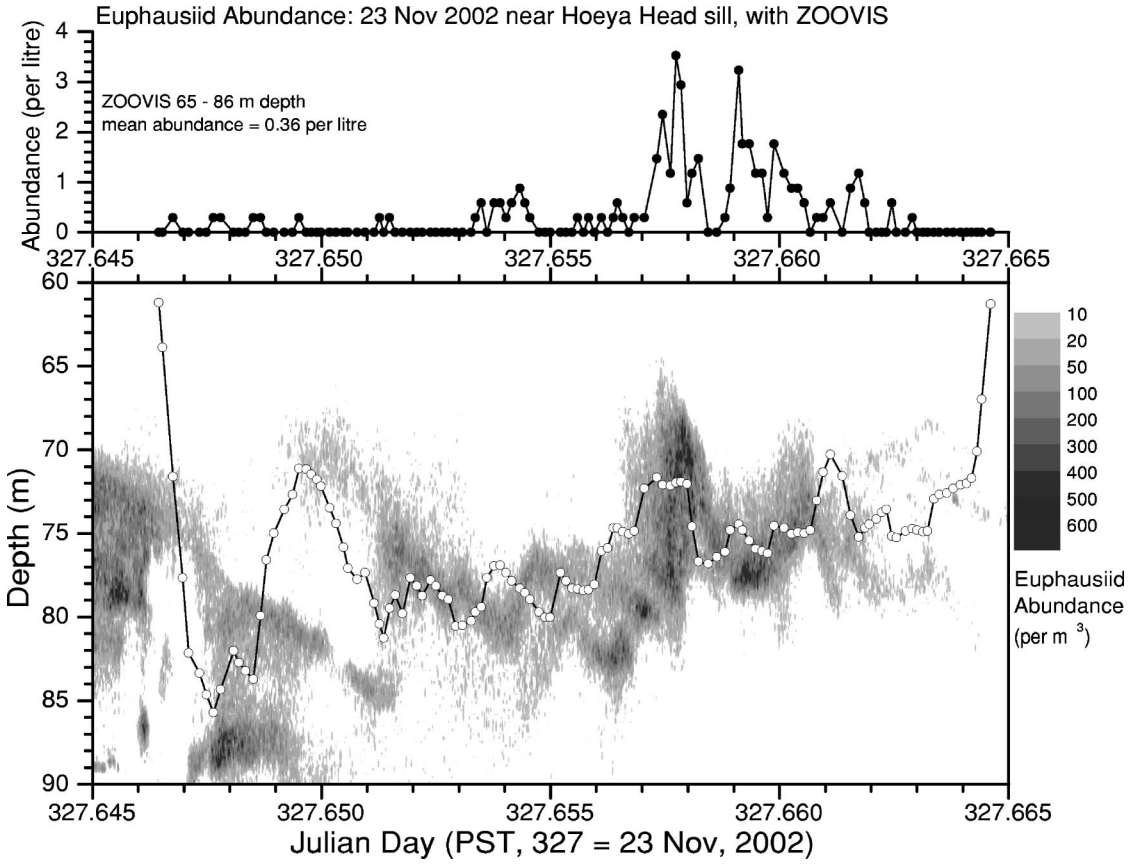


FIG. 11. Comparison of echo-sounder and ZOOVIS abundance time-series, 1528h to 1557h 23 Nov. 2002 near Hoeya Head sill. Upper panel shows ZOOVIS abundances (per liter) based on a 3.41-liter sampling volume. Lower panel shows echo-sounder abundance (per m^3) vs depth and time calculated from 120-kHz echo-sounder assuming 15.9-mm-length euphausiids as described in Sec. IV A. Echo-sounder data have 2-s time and 24-cm depth resolution. ZOOVIS trajectory overlain on echo-sounder contour plot, with circles indicating individual images.

strength profiles averaged over 1-min intervals immediately before and after the lights were turned on. In this example there were two scattering layers, one at 6–9 m and the other at 17–26-m depth. With the lights off, the averaged scattering strength within layer 2 is consistent with euphausiids (as described above) with average density of 6.6 per m^3. When the lights were turned on there was a clear decrease in S_v in layer 2. In the shallower portion of layer 2 (at 17–21-m depth) the depth- and ping-averaged scattering strength decreased by 9.8 and 10.8 dB at 120 and 200 kHz, respectively. Since this decrease was sudden (occurring within a few seconds), the euphausiids could not have had time to swim down, but rather must have simply changed their swimming-angle orientation. Since this magnitude is much larger than can be accounted for by going from a near-horizontal to a uniformly distributed swimming angle (which is about 5 dB; see Fig. 5), the most likely explanation is that the euphausiids have turned from near-dorsal to near-vertical incidence (either head or tail up). Since there is a suggestion in the *Lights On* profile that some part of scattering layer 2 has migrated downward to almost 35 m, a likely explanation is that these euphausiids have turned to swim downwards. Sameoto (1993) observed similar euphausiid photosensitivity to strobe lights mounted on a BIONESS. Interestingly, the zooplankton scattering strength in layer 1, much nearer the ship and the surface, actually increased by approximately 1.5 dB during the same period.

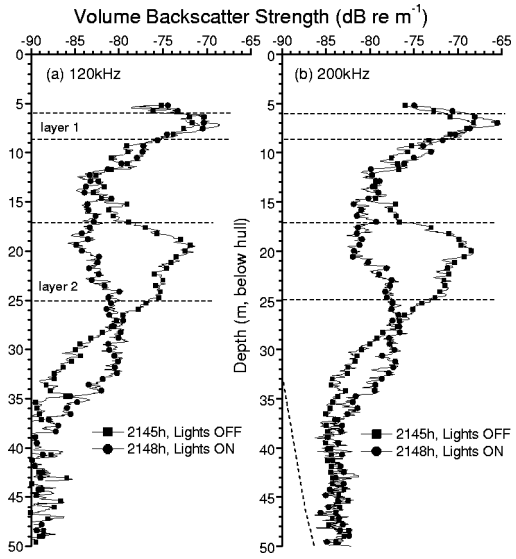


FIG. 12. Comparison of S_v vs depth profiles at (a) 120 and (b) 200 kHz, showing effects of light on euphausiids at night. Each profile averaged over 60 pings (1 min) starting at 2145h and 2148h on 21 Nov. 2002 near the sill at Hoeya Head. Depth-varying dashed line in (b) is the 200-kHz systemic noise level. Horizontal dashed lines identify two distinct scattering layers.

V. DISCUSSIONS AND CONCLUSIONS

Overall, this study showed the necessity of combining multifrequency acoustic backscatter with *in situ* net and op-

tical samples. The echo-sounders offered greater coverage rate and relatively high resolution, particularly in the fact that they sampled surface-to-seabed nearly instantaneously. The strength of the *in situ* samplers was clearly the ability to identify species and quantify size and average abundance. What emerged is an approach where the echo-sounders were employed as a spatial and temporal bridge between the isolated *in situ* samples. In these surveys the echo-sounder frequencies of 38, 120, and 200 kHz provided good coverage of the acoustically relevant portion of the zooplankton scattering spectrum, while at the same time providing reasonable sensitivity to depths of 100 to 200 m. A clear requirement for the echo-sounders was an acoustic calibration, specifically including on-axis sensitivity, beam pattern, time-varying gain curves, and instrumental noise limits.

By closely coupling the acoustic vs BIONESS comparisons through relatively homogeneous and mono-specific zooplankton scattering layers, acoustic scattering models for crustacean zooplankton and siphonophores could be validated. In particular, the echo-sounder observations of euphausiid and amphipod scattering layers satisfied a number of assumptions in the averaged fluid-cylinder model, namely the superposition of echoes from a large number of animals, a Gaussian-distributed size distribution (e.g., Fig. 3), and the dominance of a single scatterer type. The comparisons were aided by the fact that the BIONESS and acoustic sample volumes (roughly 50 to 100 m^3) were approximately matched. Furthermore, the ability to steer BIONESS to specific depths, allowing samples of particular scattering layers, greatly facilitated the comparisons. Overall, for crustacean scattering the agreement between observed and modeled TS at 120 and 200 kHz was excellent. Additionally, averaged TS and swimming orientation estimates were largely consistent with those previously reported in the literature. Unfortunately, in most situations encountered in this study the zooplankter abundances were too low to enable ZOOVIS to adequately sample the *in situ* orientation distributions. It was found that measurements at 38 kHz were occasionally contaminated by nonzooplankton echoes, particularly for cases of low euphausiid abundances and for the amphipod scattering. This contamination was readily identified in the S_v profiles as a spiky, irregular depth variation as opposed to the relatively smooth profiles exhibited by the 120- and 200-kHz channels (i.e., the echo-amplitude statistics would be different). This contamination appeared for regions with observed S_v less than roughly -90 dB (*re*: m^{-1}). The use of an upper threshold in echo-averaging largely removed this contamination.

The euphausiid and amphipod scattering examples were best modeled using a mean swimming angle of zero (horizontal), with a standard deviations of 30° and 60°, respectively. The euphausiid swimming-angle distribution was similar to some previously reported distributions for nonmigratory, *in situ* euphausiid swarms. Kristensen and Dahlen (1986) suggested that during vertical migrations the euphausiids varied their mean swimming angle by $\pm 10°$, but that a zero mean angle and 30° standard deviation was appropriate for modeling purposes. The *in situ* measurements on Antarctic krill reported by Kils (1981), which predicted a mean swimming angle of 45° (head up), and those of Chu *et al.* (1993), which predicted mean swimming angles near 20°, may not be applicable simply due to the different species and environment. In the example discussed in Sec. IV A, the euphausiids were clearly in a deep daytime, stationary layer. It is quite possible that they could have a different mean orientation during migration events or nocturnal foraging. Additionally, Chu *et al.* suggest that inversion of multifrequency data for the *in situ* swimming orientation might be possible. However, in most of the cases examined herein insufficient zooplankton samples were collected with ZOOVIS to establish reliable swimming orientation distributions.

Two examples were presented where the acoustic effects of zooplankter orientation were thought to be significant. In a nighttime near-surface survey the presence or absence of ship deck lights produced near-instantaneous changes in volume scatter strength by up to 15 dB. Since this change occurred within a few seconds, the euphausiids could not have had time to actively swim away, but rather must have simply changed their swimming-angle orientation. This change in volume scattering strength was significantly larger than predicted by a change of orientation from near-horizontal to uniformly distributed, and indicates a probable shift to near-tail incidence. This change was also larger than could be accounted for by the $\pm 10°$ variation in mean swimming angle suggested by Kristensen and Dahlen (1986). Unfortunately, the averaged fluid-cylinder model proposed by Stanton and Chu (2000) is only valid near lateral incidence. There is a clear need for these models to be extended for head or tail incidences. Another example showing the influence of zooplankter orientation comes from the acoustic vs ZOOVIS comparisons for the dense euphausiid layers near the turbulent flows at the Hoeya Head sill. The acoustically derived abundances were consistently about a factor of 2 lower than average abundances extracted from ZOOVIS. The can be at least partially explained by a change in swimming orientation distribution, from near-horizontal to more uniformly distributed. An investigation of euphausiid orientation distributions gleaned from ZOOVIS images in the higher-abundance regions is clearly called for.

The observations of siphonophore scattering in the upper inlet were unexpected. The siphonophore scattering signature exhibited a *reversed* scattering strength vs frequency, i.e., with the 38-kHz volume scattering the highest and 200 kHz the lowest. Because of the fragility of these gelatinous colonies, only isolated nectophores, pneumatophores, and siphonulae were recovered with BIONESS. An acoustic technique based on echo-amplitude statistics was used to extract siphonophore abundances and TS, producing reasonable estimates. The acoustically derived siphonophore abundance (near 0.15 per m^3) was consistent with the net samples. Similar to previous studies, the siphonophore scattering was found to be consistent with a damped bubble scattering model. The observed TS at 120 and 200 kHz was best fit by a 0.4-mm-diameter bubble size, similar to pneumatophore diameters reported by other investigators. However, at frequencies near 38 kHz the simple bubble model showed a strong resonance, generally overpredicting the observed TS by approximately 5 to 10 dB. Given the fact that the modeled

resonant peak exceeds the high-frequency TS by more than 30 dB, this discrepancy is not unreasonable. However, a suggestion for enhanced viscous damping by the pneumatophore membrane could be seen from two features in the 38-kHz scattering signature: (i) the apparent agreement at 38 kHz with a modified bubble model with 1000 times greater viscous damping, and (ii) the lack of a resonant peak in the 38-kHz S_v profile near 45 m depth. Clearly, this issue of pneumatophore damping would benefit from a careful laboratory investigation focused on the near-resonance region.

An interesting feature of echo-sounder vs ZOOVIS comparisons was the large difference in the sampling volumes of the two sensors (roughly a factor of 10^4). In its present configuration, the sampling rate and resolution of the camera were limited to 0.125–0.2 Hz and sampling volumes of 0.3 to 3.4 liters. In this work, ZOOVIS estimates of euphausiid abundances generally exceeded acoustically derived estimates by a factor of 2–4. In addition to issues surrounding zooplankter orientation discussed above, we believe that this mismatch in the sampling volumes potentially explains some of this difference. To test the effects of sample volume, a simple numerical model was created. In this model, a randomized distribution of a large number of particles (10^5) with known average density was generated. Then, a test sample volume was moved through this field of particles and the numerical density was calculated over a limited number of trials (500), simulating multiple echo-sounder pings or ZOOVIS images. The relevant parameter here is the size of the sample volume relative to the *critical volume*, defined as the reciprocal of the average particle density. For larger, supercritical sample volumes the distribution of estimated particle densities was narrowly distributed around the true average density, with the distributions increasingly narrow with larger sampling volumes. However, for very small sampling volumes ($<1\%$ of the critical volume), the distribution of estimated particle densities was bimodal, with a large number of zero estimates and a significant number of large overestimates. For a limited number of trials in the small sample volume case (i.e., modeling the ZOOVIS data collection), the overall mean density was found to be biased high by up to a factor of 3. This simple modeling suggests then that estimates of mean particle density will depend on the relative criticality of the sample volume and the number of samples. Additionally, natural zooplankton layers generally exhibit spatial *patchiness*, and thus some dependence on sampling volume size relative to patch size should be expected. Clearly, more work on the statistics of small volume samplers should be performed.

ACKNOWLEDGMENTS

This work was supported by grants from the US Office of Naval Research, Biological and Chemical Oceanography (Dr. Jim Eckman). BIONESS surveys and postanalyses were ably handled by Doug Yelland, Moira Galbraith, and Doug Moore, all of the Institute of Ocean Sciences. Finally, the authors are grateful for the skill and hard work from the officers and crew of the *CCGS* VECTOR during the two field trips in 2001 and 2002.

Benfield, M., Schwehm, C., Fredericks, R., Squyres, G., Keenan, S., and Trevorrow, M. (**2003a**). "ZOOVIS: A high-resolution digital still camera system for measurement of fine-scale zooplankton distributions," in *Scales in Aquatic Ecology: Measurement, Analysis and Simulation*, edited by P. Strutton and L. Seuront (CRC Press, Boca Raton, FL).

Benfield, M., Lavery, A., Wiebe, P., Greene, C., Stanton, T., and Copley, N. (**2003b**). "Distributions of physonect siphonulae in the Gulf of Maine and their potential as important sources of acoustic scattering," Can. J. Fish. Aquat. Sci. **60**, 0759–772.

Chu, D., Foote, K., and Stanton, T. (**1993**). "Further analysis of target strength measurements of Antarctic krill at 38 and 120 kHz: Comparison with deformed cylinder model and inference of orientation distribution," J. Acoust. Soc. Am. **93(5)**, 2985–2988.

Farmer, D., and Armi, L. (**1999a**). "The generation and trapping of solitary waves over topography," Science **283**, 188–190.

Farmer, D., and Armi, L. (**1999b**). "Stratified flow over topography: The role of small-scale entrainment and mixing in flow establishment," Proc. R. Soc. London, Ser. A **455**, 3221–3258.

Foote, K., Everson, I., Watkins, J., and Bone, D. (**1990**). "Target strengths of Antarctic krill (*Euphausia superba*) at 38 and 120 kHz," J. Acoust. Soc. Am. **87(1)**, 16–24.

Francois, R., and Garrison, G. (**1982**). "Sound absorption based on ocean measurements II. Boric acid contribution and equation for total absorption," J. Acoust. Soc. Am. **72**, 1879–1890.

Greene, C., Wiebe, P., and Burczynski, J. (**1989**). "Analyzing zooplankton size distributions using high-frequency sound," Limnol. Oceanogr. **34(1)**, 129–139.

Greene, C., Wiebe, P., Pershing, A., Gal, G., Popp, J., Copley, N., Austin, T., Bradley, A., Goldborough, R., Dawson, J., Hendershott, R., and Kaartvedt, S. (**1998**). "Assessing the distribution and abundance of zooplankton: A comparison of acoustic and net-sampling methods with D-BAD MOCNESS," Deep-Sea Res., Part II **45**, 1219–1237.

Greenlaw, C. (**1979**). "Acoustical estimation of zooplankton populations," Limnol. Oceanogr. **24(2)**, 226–242.

Harrison, P., Fulton, J., Taylor, F., and Parsons, T. (**1983**). "Review of the biological oceanography of the Strait of Georgia: Pelagic environment," Can. J. Fish. Aquat. Sci. **40**, 1064–1094.

Holliday, D. V., and Pieper, R. (**1980**). "Volume scattering strengths and zooplankton distributions at acoustic frequencies between 0.5 and 3 MHz," J. Acoust. Soc. Am. **67(1)**, 135–146.

Holliday, D. V., and Pieper, R. (**1993**). "Bioacoustical oceanography at high frequencies," ICES J. Mar. Sci. **52**, 279–296.

Johnson, R. (**1977**). "Sound scattering from a fluid sphere revisited," J. Acoust. Soc. Am. **61**, 375–377.

Kils, U. (**1981**). "The swimming behavior, swimming performance, and energy balance of Antarctic krill, *Euphausia superba*," BIOMASS Sci. Ser. **3**, 122 pp.

Klymak, J., and Gregg, M. (**2001**). "Three-dimensional nature of flow near a sill," J. Geophys. Res. **106(C10)**, 22295–22311.

Klymak, J., and Gregg, M. (**2003**). "The role of upstream waves and downstream density pool in the growth of lee waves: Stratified flow over the Knight Inlet sill," J. Phys. Oceanogr. **33**, 1446–1461.

Klymak, J., and Gregg, M. (**2004**). "Tidally generated turbulence over the Knight Inlet sill," J. Phys. Oceanogr. **34(5)**, 1135–1151.

Kristensen, A., and Dahlen, J. (**1986**). "Acoustic estimation of size distributions and abundance of zooplankton," J. Acoust. Soc. Am. **80(2)**, 601–611.

Korneliussen, R. (**2000**). "Measurement and removal of echo integration noise," ICES J. Mar. Sci. **57(4)**, 1204–1217.

Mackie, G., and Mills, C. (**1983**). "Use of the Pisces IV submersible for zooplankton studies in coastal waters of British Columbia," Can. J. Fish. Aquat. Sci. **40**, 763–776.

Mackie, G. (**1985**). "Midwater macroplankton of British Columbia studies by submersible PISCES IV," J. Plankton Res. **7(6)**, 753–777.

Medwin, H., and Clay, C. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, San Diego).

Miyashita, K., Aoki, I., and Inagaki, T. (**1996**). "Swimming behavior and target strength of Isada krill (*Euphausia pacifica*)," ICES J. Mar. Sci. **53**, 303–308.

Robison, B., Reisenbichler, K., Sherlock, R., Silguero, J., and Chavez, F. (**1998**). "Seasonal abundance of the siphonophore *Nanomia bijuga* in Monterey Bay," Deep-Sea Res., Part II **45**, 1741–1751.

Rogers, C. (**1978**). "Aggregation of the siphonophore *Nanomia cara* in the

Gulf of Maine: Observations from a submersible,'' Fish. Bull. **76**, 281–284.
Ross, T., and Lueck, R. (**2003**). ''Sound scattering from oceanic turbulence,'' Geophys. Res. Lett. **30**, 1344–1347.
Sameoto, D., Jarozynski, L., and Fraser, W. (**1980**). ''BIONESS: A new design in multiple net zooplankton samplers,'' Can. J. Fish. Aquat. Sci. **37**, 722–724.
Sameoto, D., Cochrane, N., and Herman, A. (**1993**). ''Convergence of acoustic, optical, and net-catch estimates of euphausiid abundance: use of artificial light to reduce net avoidance,'' Can. J. Fish. Aquat. Sci. **50(2)**, 334–346.
Stanton, T. (**1985**). ''Density estimates of biological sound scatterers using sonar echo peak PDFs,'' J. Acoust. Soc. Am. **78(5)**, 1868–1873.
Stanton, T. (**1989**). ''Simple approximate formulas for backscattering of sound by spherical and elongated objects,'' J. Acoust. Soc. Am. **86(4)**, 1499–1510.
Stanton, T., Chu, D., Wiebe, P., and Clay, C. (**1993**). ''Average echoes from randomly oriented random-length finite cylinders: Zooplankton models,'' J. Acoust. Soc. Am. **94(6)**, 3463–3472.
Stanton, T., Wiebe, P., Chu, D., and Goodman, L. (**1994**). ''Acoustic characterization and discrimination of marine zooplankton and turbulence,'' ICES J. Mar. Sci. **51**, 469–479.
Stanton, T., Chu, D., Wiebe, P., Martin, L., and Eastwood, R. (**1998**). ''Sound scattering by several zooplankton groups. I. Experimental determination of dominant scattering mechanisms,'' J. Acoust. Soc. Am. **103(1)**, 225–235.
Stanton, T., and Chu, D. (**2000**). ''Review and recommendations for the modeling of acoustic scattering by fluid-like elongated zooplankton: Euphausiids and copepods,'' ICES J. Mar. Sci. **57**, 793–807.
Trevorrow, M., and Tanaka, Y. (**1997**). ''Acoustic and *in situ* measurements of freshwater amphipods (*Jesogammarus annandalei*) in Lake Biwa, Japan,'' Limnol. Oceanogr. **42(1)**, 121–132.
Trevorrow, M. (**1998**). ''Observations of internal solitary waves near the Oregon coast with an inverted echo-sounder,'' J. Geophys. Res. **103(C4)**, 7671–7680.
Vagle, S., Foote, K., Trevorrow, M., and Farmer, D. (**1996**). ''Absolute calibrations of monostatic echosounder systems for bubble counting,'' IEEE J. Ocean. Eng. **21(3)**, 298–305.
Warren, J., Stanton, T., Benfield, M., Wiebe, P., Chu, D., and Sutor, M. (**2001**). ''*In situ* measurements of acoustic target strengths of gas-bearing siphonophores,'' ICES J. Mar. Sci. **58**, 740–749.
Watkins, J., and Brierley, A. (**1996**). ''A post-processing technique to remove background noise from echo integration data,'' ICES J. Mar. Sci. **53**, 339–344.
Wiebe, P., Greene, C., Stanton, T., and Burczynski, J. (**1990**). ''Sound scattering by live zooplankton and micronekton: Empirical studies with a dual beam acoustic system,'' J. Acoust. Soc. Am. **88(5)**, 2346–2360.
Wiebe, P., Stanton, T., Benfield, M., Mountain, D., and Greene, C. (**1997**). ''High-frequency acoustic volume backscattering in the Georges Bank coastal region and its interpretation using scattering models,'' IEEE J. Ocean. Eng. **22(3)**, 445–464.

The significance of high-order resonances of spherical bubbles to the acoustic response of fish with swimbladders

Richard Barr[a] and Roger F. Coombs[b]
National Institute of Water and Atmospheric Research Ltd., P.O. Box 14-901, Kilbirnie, Wellington, New Zealand

(Received 7 March 2003; accepted for publication 22 March 2005)

The high-order resonant response of spherical gas bubbles in seawater, when insonified by an acoustic pulse of 12 cycles of 38 kHz, used for target strength measurements, shows substantial variation in shape with only small changes in bubble radius. In contrast, the amplitude response of a gas bubble near its fundamental resonance changes predominantly in magnitude as the radius changes. The fundamental response is very similar to that of a more rigid target such as a tungsten carbide sphere. It might initially be thought that the high-order resonant response of spherical gas bubbles in seawater would have little relevance to the response of a gas-filled spheroidal swimbladder immersed in a viscous fish body. However, target responses similar to those predicted theoretically for gas bubbles have been found in data collected *in situ* for estimating target strength from normal fish populations. With further theoretical and experimental work, high-order resonances could be a useful aid to estimating target size and possibly target species. [DOI: 10.1121/1.1905638]

PACS numbers: 43.30.Sf, 43.20.Fn, 43.80.Ev [KFG] Pages: 3589–3599

I. INTRODUCTION

In this paper we investigate the significance of high-order resonances in gas-filled bladders in fish and other marine organisms in identifying fish targets and estimating fish target strengths by means of a simple resonance model and *in situ* data from the ocean.

It is now more than a century since Lord Rayleigh (1896) investigated acoustic scattering by fluid spheres with dimensions very small compared to the wavelength and more than 50 years since Anderson (1950) advanced the study to encompass the scattering of sound from fluid spheres with diameters less than a few wavelengths. These works have been extended to derive the resonant response of gas bubbles in water (Medwin and Clay, 1998), the resonant backscatter from plankton that contain gas inclusions, such as siphonophores (Stanton and Chu, 1998; Warren *et al.*, 2001), and fish containing gas-filled swimbladders (Love, 1978). The resonant response of a fish swimbladder is reduced in comparison to that of an equivalent free gas bubble and exhibits a lower Q value, due to the damping effects produced by the viscosity of the fish flesh surrounding the gas-filled bladder (Love, 1978). Numerous papers have been written extending the spherical analysis to the more general case of a prolate spheroid (Furusawa, 1988; Silbiger, 1963; Yeh, 1967), but this increases the numerical complexity considerably.

Most experimental measurements of swimbladder resonance have tended to concentrate on the fundamental, omnidirectional, breathing-mode resonance (Batzler and Pickwell, 1970; McCartney and Stubbs, 1970; Sand and Hawkins, 1973) and the weaker higher-order resonances, as presented, for example, in Anderson (1950), have largely been ignored. This is understandable, as most previous work has been directed towards estimation of target strength for use in stock assessment where the strongest scatterers are the most important. However, our primary motivation was the identification of single fish targets and to this end we compute the resonant response of small spherical gas bladders to insonification with short sound pulses where the sphere radius is less than a wavelength. For *in situ* target strength and target identification studies we typically use a 12-cycle pulse (a ''ping'') of 38 kHz (Macaulay, 1997; McClatchie and Coombs, 2000). This has a duration of about 316 μs and a scale length in sea water of about 47 cm and is the shortest practical stable pulse that is usable with our system. This short pulse maximizes the chance of detecting single target echoes and also brings back information from the target over a wide frequency range due to its broad bandwidth. Computed pulse responses can be compared directly with the *in situ* single fish target responses.

We show how the form of the acoustic ping response of single spherical gas bubble targets can be significantly affected by the high-order resonances and, qualitatively, how the myriad of different ping responses of resonant gas bubble targets are mirrored in the experimentally observed ping responses of swimbladdered fish swimming freely in the ocean. The application of simple gas bubble theory to fish was previously considered impractical because it neglected damping by the viscous fish body (Johnson, 1977). However, this study shows that spherical gas bubble theory can provide a significant insight into the acoustic ping response of swimbladdered fish.

[a]Corresponding author. Present address: Blue Mountains Road, Whitemans Valley, Wellington, New Zealand.
Electronic mail: r.barr@clear.net.nz
[b]Present address: Innovative Solutions Limited, P.O. Box 22-235, Wellington, New Zealand.

II. MODELING

To compute the acoustic ping response at a frequency, f, of a single spherical gas bubble, it is first necessary to calculate the complex target strength, $S(f)$, of a spherical gas bubble in seawater. To do this we follow the work of Anderson (1950) as summarized by Johnson (1977) or Medwin and Clay (1998). If $P(f)$ is the frequency spectrum of the 12-cycle ping of 38 kHz insonifying this gas bubble, then the total system response, $R(f)$, to the ping can be evaluated using Eq. (1),

$$R(f)=P(f)S(f)M(f), \tag{1}$$

assuming we know the frequency response, $M(f)$, of the acoustic recording system. To simulate the response of the acoustic system, $M(f)$, we used transmit and receive transfer functions based on standard equivalent circuits (e.g., Stansfield, 1991) of the form

$$T_{t,r}(f)=[1-\tfrac{1}{2}Q_mQ_{t,r}X^2+\tfrac{1}{2}i(Q_m+Q_{t,r})X]^{-1}, \tag{2}$$

where $X=(f/f_0-f_0/f)$ and Q_m, Q_t, and Q_r are the mechanical, transmitter, and receiver Q's respectively, and $M(f)=T_t(f)T_r(f)$. We took $Q_m=8.1$, $Q_t=1.27$, and $Q_r=10.77$ from MacLennan (1981), which are for a Simrad transducer similar to the one we used for data collection (see Sec. III). Frequency response information was not supplied with our transducer, but spot measurements we have made show that the above transfer functions provide a good description.

In the simulations, the range to the theoretical target is kept constant and all results are normalized to the theoretical response of a tungsten carbide calibration sphere (see later) at the same range. By taking the inverse Fourier transform of $R(f)$ and complex demodulating, we obtain the pulse response, $R(t)$, of the bubble to insonification by the transmitted ping which can be compared with *in situ* data as outlined earlier.

III. EQUIPMENT

All the *in situ* experimental 38-kHz ping response data presented here were collected using the National Institute of Water and Atmospheric Research's (NIWA) Computerized Research Echo Sounder Technology (CREST) (Coombs *et al.*, 2003). For this work the system consisted of a Simrad ES38DD split-beam transducer with underwater electronics mounted on an open, stainless steel frame as described in Barr *et al.* (2002). The complete system is referred to hereafter as the ''frame'' and it was deployed from NIWA's research vessel *Tangaroa* while slowly drifting. For target strength and target identification measurements the system transmitted a sinusoidal pulse of 12 cycles of 38 kHz at 1.4-s intervals. The received data were filtered, complex demodulated, and corrected for spherical spreading and sound absorption with a time varied gain. The filtered *in situ* data were then stored for later analysis in the New Zealand Ministry of Fisheries fisheries acoustics database, which is managed by NIWA. In this database, entries are characterized by a voyage identifier (e.g., ''TAN0011''), whether the data were collected by a towed vehicle (''towbody'') or other means, and by a file identifier (e.g., ''d1,'' ''d2,'' etc.).

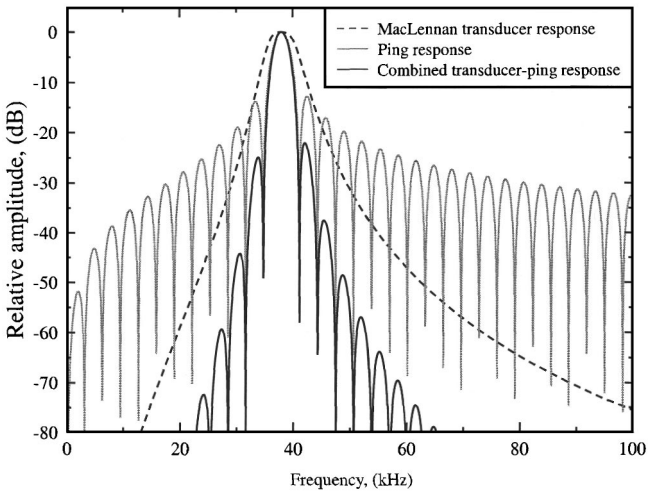


FIG. 1. Spectral amplitude of a 12-cycle ping plotted together with, and combined with, the MacLennan (1981) transducer frequency response.

The data handling capacity of the system imposes a range limit that for the target strength mode was about 100 m. The frame was always deployed with a tungsten carbide calibration sphere of radius 19.05 mm suspended about 10 m beneath it [in a similar fashion to Kloser *et al.* (2000)]. In the prevailing oceanographic conditions the sphere had a target strength of −42.34 dB at 38 kHz and was visible in most transects, allowing direct comparison of fish and sphere target strengths.

As shown in Eq. (1) the ping response of a target depends on the spectrum of the transmitted pulse, $P(f)$, and the bandwidth of the acoustic recording system, $M(f)$, and these are plotted in Fig. 1. The ping spectrum is computed for a pulse of 12 cycles of a 38-kHz sinusoid, as used in our experimental recordings, and is presented as the continuous gray line in Fig. 1. It can be seen that between 20 and 90 kHz the pulse spectrum is within 30 dB of the peak response (normalized to 0 dB at 38 kHz). However the bandwidth at the −12-dB level is only about 10 kHz. The system response, $M(f)$, is shown by the chained line in Fig. 1. The system bandwidth at the −12-dB level is similar to the spectrum of the transmitted ping. The solid black line in Fig. 1 represents the combined spectrum of the transmit-receive system and the transmitted pulse spectrum. For a broadband scatterer the black trace thus closely approximates the spectrum of the ping recorded by CREST. Note the spectrum only has a bandwidth of 5 kHz at the −20-dB level. This indicates that our acoustic system will only respond to very localized features in the frequency spectrum of any bubbles being insonified.

IV. COMPUTATIONS

Using the model from Sec. II, we have computed the complex spectral response, $S(f)$, of a series of spherical bubbles of gradually increasing radius. To give the results a fisheries perspective, we have normalized the volume of the spherical bubbles to the volume of the swimbladder of a

TABLE I. Physical parameters for spheroidal model myctophid, total length 7.5 cm.

Parameter	Units	Swimbladder	Sea water
Length	cm	1.5	n/a
Height	cm	0.5	n/a
Width	cm	0.5	n/a
Velocity of sound	m/s	335.0	1484
Density at 800 m	kg/m^3	100	1032

7.5-cm myctophid (*Myctophum punctatum*) as described by Marshall (1960). The dimensions, scaled from his figure, are presented in Table I together with data on the density of sea water and swimbladder gas applicable to a depth of 800 m. This is the depth, on the Chatham Rise, at which most of our experimental data were recorded. Our reference ''myctophid'' thus has a scale size of 100% corresponding to a spherical bubble of radius 3.6 mm. Computations are presented up to a scale range of 675%, an effective swimbladder (or bubble) radius of about 5 cm which should approximate the response of much larger fishes such as black and smooth oreo (*Allocyttus niger* and *Pseudocyttus maculates*).

The adoption of a spherical rather than a spheroidal shape for the swimbladder will modify both the frequencies and *Q*'s of the resonances observed, but for the small aspect ratio of the myctophid swimbladder (3:1), the effects should not be too great (Feuillade and Werby, 1994). It should also be remembered that the purpose of this paper is to investigate the characteristics of the ping signatures of high-order resonances of swimbladders rather than to obtain quantitative values of target strength or the frequency of resonance of a specific type and size of fish.

Figure 2 shows the spectral amplitude and temporal amplitude responses of three spheres with radii 27%, 21%, and 17% that of the reference myctophid. The gray column centered on 38 kHz in the upper panel indicates the bandwidth of the combined ping and transducer system at the −20-dB level (see Sec. III). As expected, when the sphere resonance falls within the system bandwidth, the amplitude of the total response is greatest. The response plots have been displayed with abscissas scaled in ''range (m)'' to be compatible with the *in situ* target strength from the Chatham Rise that will be presented later. By dividing the abscissa by half the velocity of sound in seawater (see Table I) the abscissa can be converted to the more common ''elapsed time(s)'' format. It can be seen that for the fundamental resonance, the form of the ping response does not change markedly as the resonance moves through the optimum sensitivity of the system and is very similar in form to those computed for the tungsten carbide calibration sphere used to normalize the ping amplitude response scale. However, they are somewhat smaller in amplitude, the sphere response being −42.34 dB.

The format of Fig. 3 is similar to that of Fig. 2, but in this case the data are plotted for spheres with radii 75%, 88%, and 105% that of the reference myctophid. This suggests that these results might be representative of a typical real adult myctophid. The upper panel in Fig. 3 shows the amplitude spectrum. The scaling on the *y* axis applies directly to the sphere of radius 105%. It can be seen that as

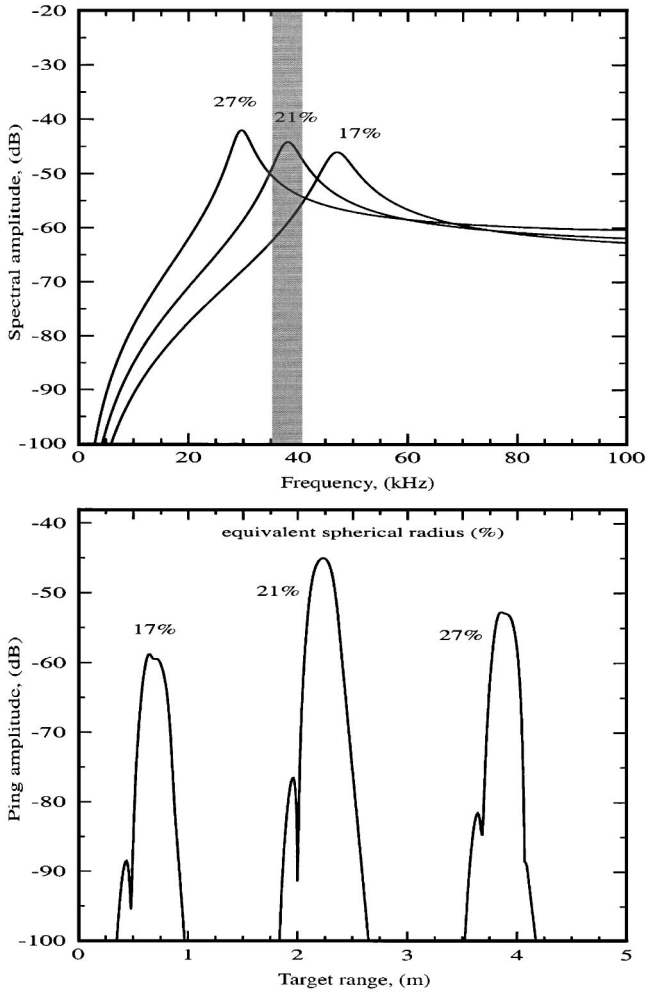


FIG. 2. Upper panel: amplitude spectra of three spherical gas bubbles near their fundamental resonance. Lower panel: ping envelope responses for the three spherical bubbles whose spectra are presented in the upper panel. Ping envelope plots have been displaced relative to each other along the *x* axis to facilitate comparison. The percentage (%) scalings give the bubble radius referenced to the radius of a sphere with the same volume as the swimbladder of a 7.5-cm myctophid (Marshall, 1960).

well as the fundamental resonance, now near 8 kHz, at least four higher-order resonances and inflections (Rudgers, 1967) are visible within the 100-kHz display range of the figure. These amplitude spectra are very similar to the spectrum given in Fig. 3 of Johnson (1977), computed for a bubble at a depth of 350 m, which lends credence to the present results. The sphere radii have been chosen so that one (88%) has its second resonance centered at the recording system's optimum frequency sensitivity (gray band in the upper panel of Fig. 3) while the other two spheres (75% and 105%) resonate just above and below the optimally sensitive region, respectively. The resulting acoustic system amplitude responses are shown in the lower panel of Fig. 3. Also note that the abscissa is plotted on a scale twice as coarse as Fig. 2. As might be expected, the spherical bladder (88%), with its second resonance at the peak of system sensitivity, produces the maximum ping response. However, what is significantly different from Fig. 2 is the long slow exponential decay in the response of this sphere. The response for the sphere of 75% radius also has a slight exponential decay, but

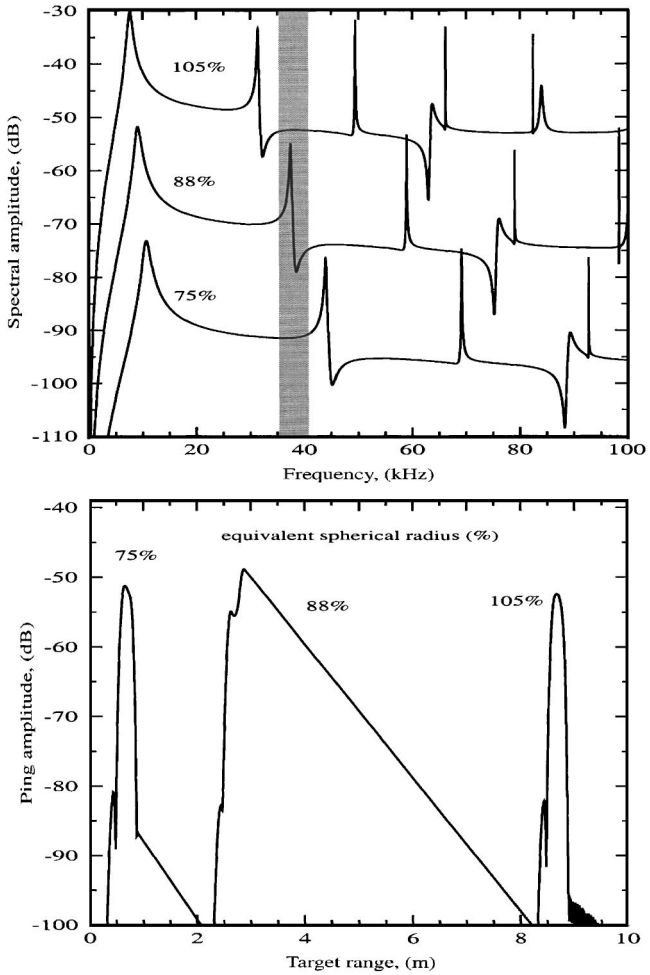


FIG. 3. Upper panel: amplitude spectra of three spherical gas bubbles near their second order resonance. The *y* axis applies directly to the top curve (105%). The *y* axis for the middle and bottom curves has been displaced by −20 and −40 dB, respectively, to facilitate shape comparison. Lower panel: ping envelope responses for the three spherical bubbles whose spectra are presented in the upper panel. See Fig. 2 caption for *x*-axis displacement and % scaling details.

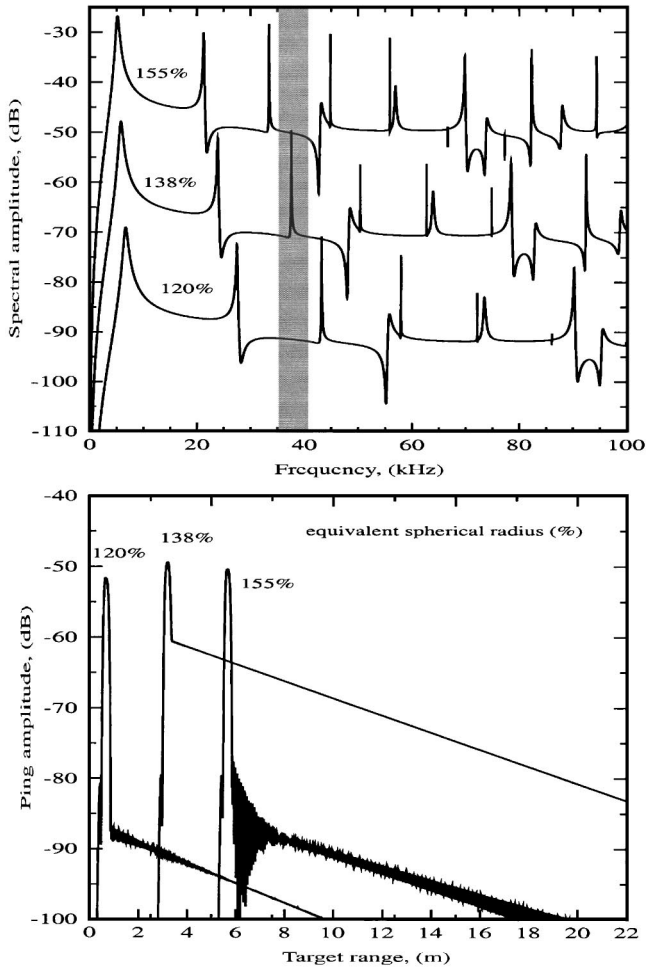


FIG. 4. Upper panel: amplitude spectra of three spherical gas bubbles near their third-order resonance. The *y* axis applies directly to the top curve (155%). The *y* axis for the middle and bottom curves has been displaced by −20 and −40 dB, respectively. Lower panel: ping envelope responses for the three spherical bubbles whose spectra are presented in the upper panel. See Fig. 2 caption for axis displacement and % scaling details.

it starts some 35 dB lower than the maximal response. The response of the sphere with a 105% radius is most similar to the responses to the fundamental resonance presented in the lower panel of Fig. 2. It thus appears that myctophidlike fish targets will give very different acoustic responses as their sizes change over a very small scale-range (75%–105%). Indeed, it was initially thought these marked differences in response resulted from different fish species (Barr *et al.*, 2002)!

In the upper panel of Fig. 4 we present the spectral response of three larger bubbles, chosen such that their third acoustic resonance again straddles the most sensitive region of the system response. The envelope of the ping response to these bubbles is shown in the lower panel of Fig. 4. Note that the horizontal axis is now more than four times coarser than that of Fig. 2. The pulse envelopes of the spheres with their third resonance outside the optimal system response are single peaked responses very similar to those presented in Figs. 2 and 3 for the bubbles resonating outside the peak system bandwidth. However, the bubble (138%) with its third resonance at the peak system response again has an exponential decay. In this case the signal drops rapidly by about 10 dB before developing the exponentially decaying characteristic. The decay is much slower than that in the lower panel of Fig. 3 and this is probably because the third resonance is much sharper than the second. Figure 4 further demonstrates that a small change in bubble radius (120%–155%) can produce a significant change in the envelope of the acoustic response.

It is clear from Figs. 2–4 that the bubble resonance has to occur near the peak of the system sensitivity to significantly change the envelope of the ping response. We will therefore only present the acoustic response of larger bubbles for the case where the higher-order resonances lie within the optimal range of the system response. The left-hand side of Fig. 5, from the bottom up, shows the amplitude spectra of the acoustic responses of three bubbles of radii 178%, 230% and 290%. For the smallest bubble (178%) the first four resonances have been labeled. It is clear that whereas the first three resonances are singlet features, the fourth is closely associated with an inflection. Indeed, for radii >170% it seems impossible to get a single resonance feature within the bandwidth of the receiving system. The right-hand side of

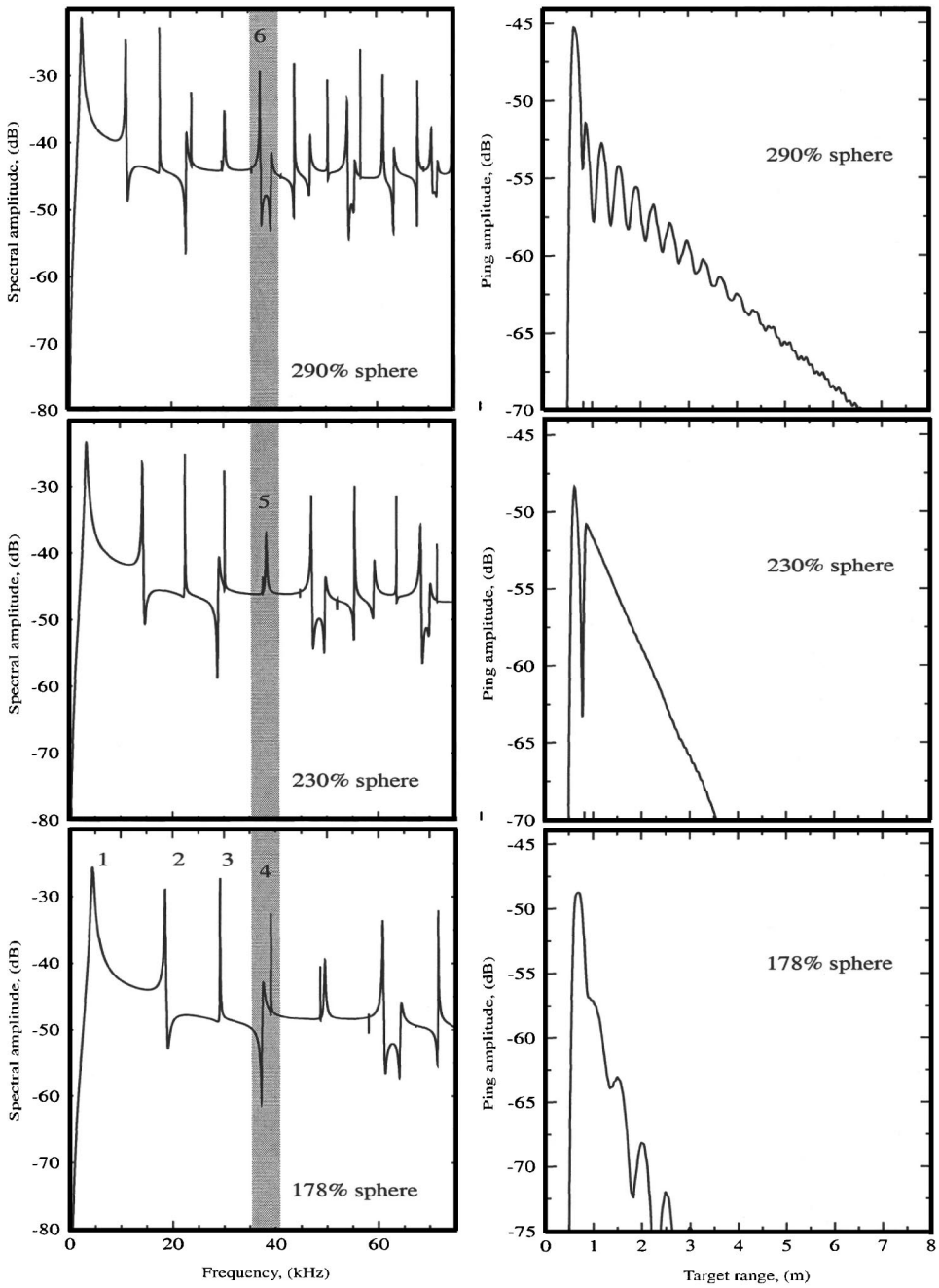


FIG. 5. Left-hand panels: amplitude spectra of three spherical gas bubbles with their fourth-, fifth-, and sixth-order resonances near 38 kHz. Right-hand panels: ping envelope response for the three spherical bubbles whose spectra are presented in the left-hand panel. See Fig. 2 caption for % scaling details.

Fig. 5 shows the envelopes of the responses for the three sphere targets whose spectra are shown on the left of the figure. What is remarkable is the great difference in shape of the pulse response of the bubble as its radius is increased. Without this computational study one would consider the different shaped ping responses resulted from much more significant changes in target structure than merely a change in scale. Finally, Figs. 6 and 7 show some typical envelope responses for spherical bubbles with radii ranging from 348% to 675% (12.5 to 25 mm). The only consistent feature in the plots is the gradual increase in the peak impulse response with increasing bubble radius. The qualitative nature of the responses changes enormously from a main peak with oscillatory decay when the radius is 348% through a single impulse response at a radius of 385%, to complicated double peaked skirt responses at radii of 630% and 675%.

V. *IN SITU* RESULTS

The *in situ* target responses presented here were all obtained from *Tangaroa* using the frame described in Sec. III. The recordings were made during a biomass survey (TAN0011) of black and smooth oreo (*Allocyttus niger* and *Pseudocyttus maculates*) on the Chatham Rise, New Zealand, in October–November 2000. Most recordings were made at a depth of ≈800 m while the vessel was slowly drifting. The slow drift had the useful byproduct that it was sometimes possible to insonify the same target many times.

The responses are shown in Figs. 8–10. Three pairs of plots are presented in each figure. The plots on the left-hand side give the ping response observed when an interesting target had entered the region insonified by the transducer and had risen above the system noise level. The file number, used

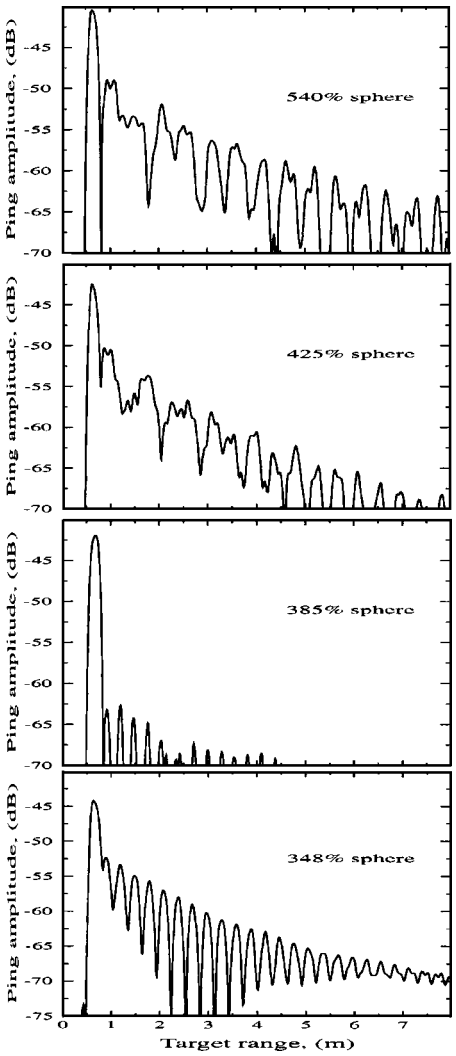


FIG. 6. Ping envelope responses for four spherical bubbles with radii between 348% and 540%. See Fig. 2 caption for % scaling details.

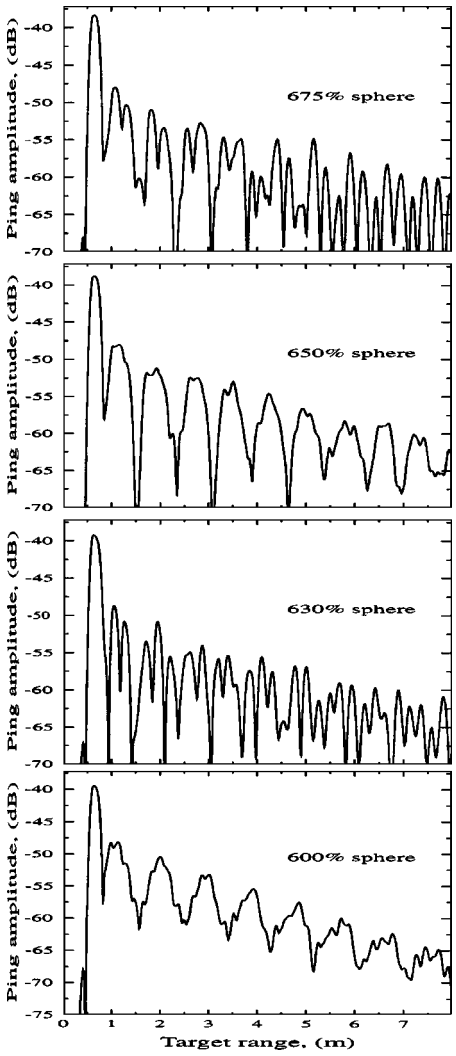


FIG. 7. Ping envelope responses for four spherical bubbles with radii between 600% and 675%. See Fig. 2 caption for % scaling details.

in database access, is given in the upper left of each plot on the left-hand side. The three plots on the right-hand side of the figure show the ping response from the same target (as shown on the left-hand side of the figure) just before it passed out of the insonified region. The ping number is given in each panel. Typically, four consecutive ''hits'' were made on a single target.

The top panel in Fig. 8 shows the response of the tungsten carbide calibration sphere, suspended about 19.5 m below the frame. The sphere was visible most of the time between pings 443 and 545. The response is less than −42.34 dB as the target sphere was not along the bore-sight of the transducer, as evinced by the upper trace in each panel. Note how the response is very similar to that of a bubble with its fundamental resonance near 38 kHz (see Fig. 2). This shows that further information (e.g., target phase) is needed to distinguish between solid and gas-filled targets in such cases. The remaining eight target sequences presented in Figs. 8–10 have been selected solely on their physical similarity to some of the target responses computed theoretically in Figs. 3–7.

Figure 8 shows the great advantage of suspending a calibration sphere below the frame during all measurements. To select a biological target or a calibration target it is only necessary to change the target range. Calibration is thus contemporaneous with data collection and is therefore subject to fewer errors than calibrations made in different circumstances, for example in a calibration tank before or after a voyage (Kloser *et al.*, 2000).

The central plot in Fig. 8 is one of three pings (2921–2923) on the same biological target that were very similar to each other but completely different to the target sphere response plotted above. In fact, the response is very similar in magnitude and shape to the response of the spherical bubble of radius 88% presented in Fig. 3. This suggests that gas bladders in fish targets respond in a very similar way to free gas bubbles and that the target was probably a small fish about the size of our reference myctophid. The upper trace in the central plot (the target bearing) shows a constant steady value during the time that the signal decays exponentially, further confirming a fixed source for the acoustic signal. This should be compared with the bearing of the target sphere plotted above. As the range to target did not decrease rapidly over the three consecutive pings, the response data were not produced from scattering by rising gas bubbles. It is unlikely

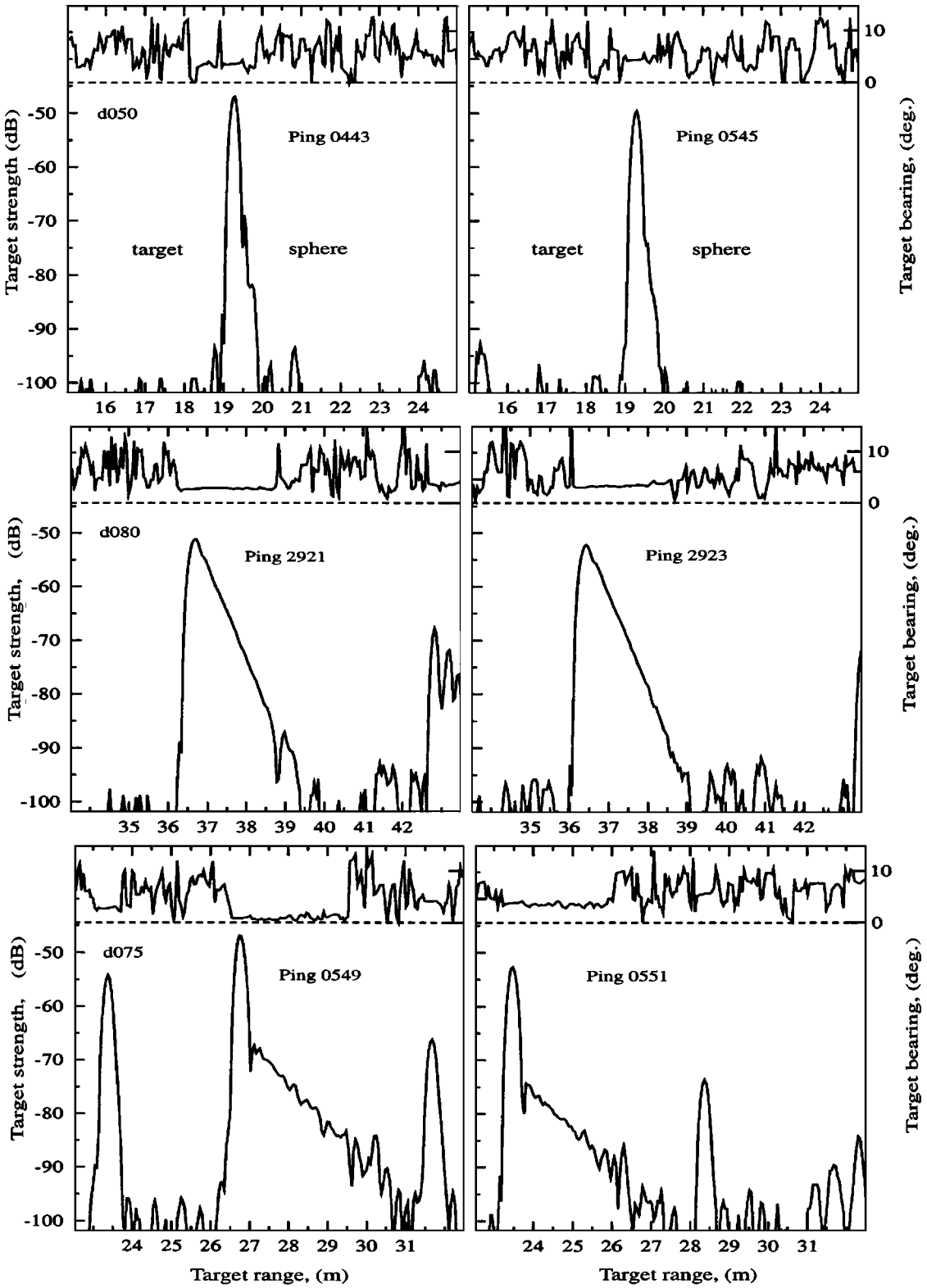


FIG. 8. Target strength versus range plots of observed *in situ* ping responses with associated bearing angle information (at the top of each figure). Left-hand panels: target strength versus range plots of the first ping in a series of similar shaped ping responses. Right-hand panels: amplitude-range plots of the last ping in the series with a similar shape to the first ping shown on the left-hand side of the figure. The two plots at the top of the figure show the response of a tungsten carbide calibration sphere suspended ≈20 m below the frame. The plots in the central and bottom panels are considered to be the responses of biological targets. The left-hand ordinate in all panels is the target strength in dB and applies to the lower trace in each panel. The right-hand ordinate is the angle to the target measured from the normal to the transducer surface and applies to the upper trace in each figure.

that a free gas bubble, of about 0.3-cm radius, would remain at a constant depth in the water column for more than 4 s.

The bottom plots in Fig. 8 show responses with clear single peaks followed by exponentially decaying skirts starting some 20 dB below the main peak. As with the central plots, these are from a series of three similar pings on the same target. The bearing data again show that all the acoustic energy originates from the same source. This response is

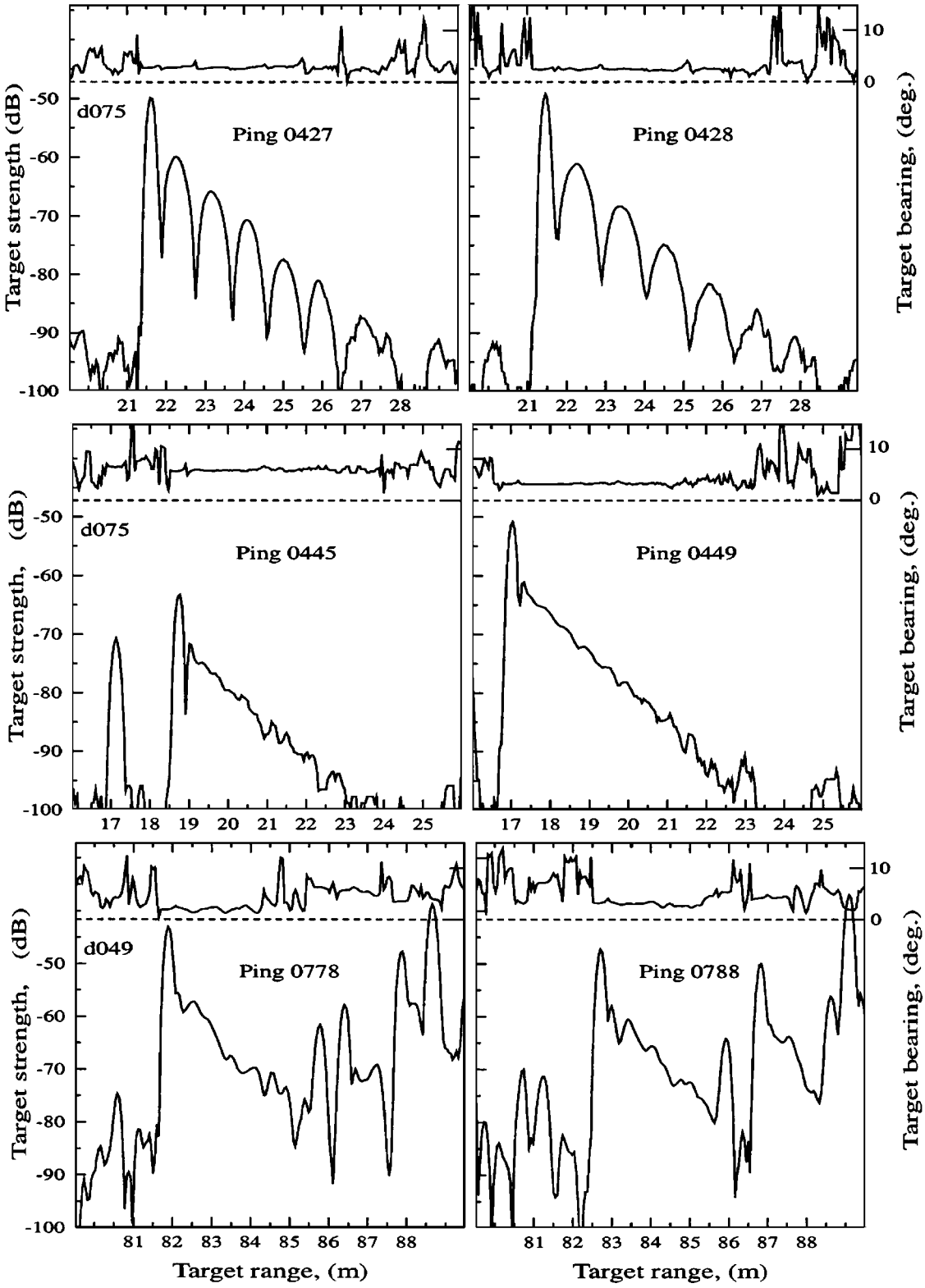


FIG. 9. Target strength versus range plots of ping sequences on three different acoustic targets. For layout details see Fig. 8.

similar in shape and magnitude to the bubble third harmonic responses shown in Fig. 4, suggesting that the fish was somewhat larger than our myctophid.

The two ping responses at the top of Fig. 9 also display a dominant initial peak, but this time the signal oscillates while also decaying exponentially. The target bearing is constant, during the main pulse and skirt decay, confirming a single target. This complex response was originally thought to result from Lamb waves circulating around some hard-cased planktonic target (Barr *et al.*, 2002), but we now consider it is more likely that it is simply produced by a high-order swimbladder resonance (see upper waveform in Fig. 5).

The central plots in Fig. 9 have a similar shape to the

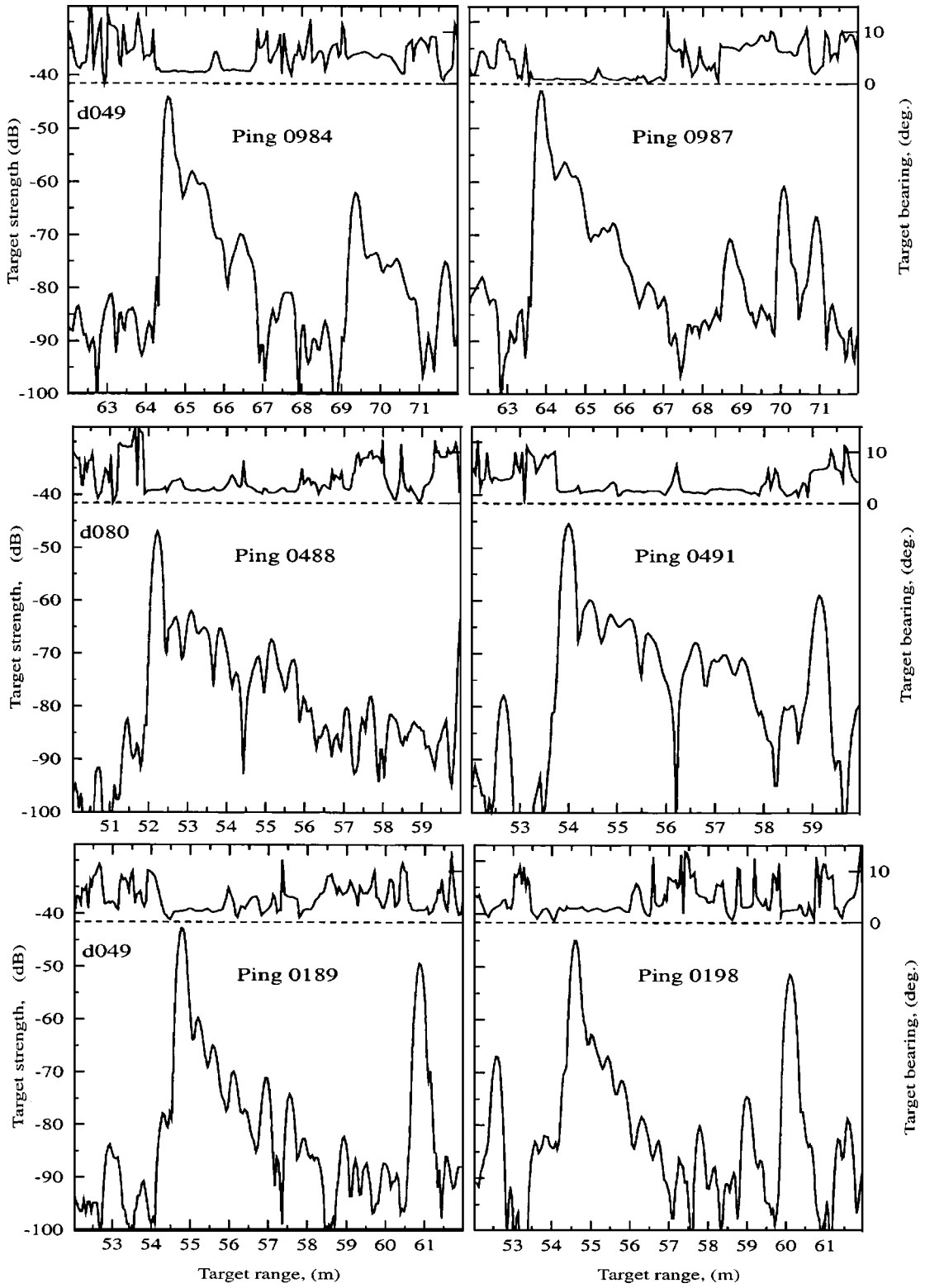


FIG. 10. Target strength versus range plots of ping sequences on three different acoustic targets. For layout details see Fig. 8.

waveform shown in the central panel of Fig. 5 and again show a consistent bearing for both the main pulse response and the exponentially decaying skirt.

The two ping responses shown at the bottom of Fig. 9 are selected from 11 similarly shaped consecutive ping responses. The skirt response, after the main peak, had a similar shape for all 11, showing that the peaks and troughs were not just the result of system noise. The almost constant target bearing, as the response decays from the peak, supports this hypothesis. Note that this target is much stronger (about −43 dB) than those plotted in Fig. 8, and is probably the resonant response of a large swimbladder (e.g., see the response of the sphere of radius 425% in Fig. 6). Such a gas bladder would have a radius of about 1.5 cm. The target remained in view at almost constant range for more than 15 s, so it is extremely unlikely that it could be a free gas bubble and we conclude

that it is the response of gas trapped in the swimbladder of a fish.

Finally, in Fig. 10, three strong target responses (≈ -45 dB) are presented. The skirt response in these figures, while complicated, is remarkably consistent from ping to ping. Those shown in the top and central panels are maintained over four consecutive pings and in the bottom panel over ten. These data show that the skirt response is consistent from ping to ping, while there are significant differences from target to target (top, center, and bottom panels). The responses in Fig. 10 also show similarities with the computed sphere responses for radii 425 to 675% in Figs. 6 and 7.

VI. DISCUSSION AND CONCLUSIONS

The amplitude (ping) response of a gas bubble in seawater, near its fundamental resonance, when insonified by an acoustic ping of 12 cycles at 38 kHz, changes predominantly in magnitude as the bubble radius changes about the resonance value. The ping response is very similar, in both magnitude and shape, to the ping response of a more rigid target such as a tungsten carbide sphere. This will cause problems with target identification if one attempts to use amplitude information only to determine the nature of acoustic targets. However, a population of gas bubble targets near fundamental resonance can be distinguished from solid targets using information on target phase (Barr and Coombs, 2004).

The amplitude (ping) response of a spherical gas bubble, near a high-order resonance, shows substantial variation in shape with only small changes in bubble radius. This shape difference appears mainly in the skirt and is best seen by plotting the ping response on a logarithmic scale. It might initially be thought that the high-order resonant response of spherical gas bubbles in seawater would have little relevance to the response of a gas-filled spheroidal swimbladder immersed in a viscous fish body (Johnson, 1977). However, we have found target responses similar to those predicted theoretically for gas bubbles in data collected *in situ* for estimating target strength from normal fish populations. The range to these targets, some at depth of about 800 m, remains almost constant for periods of more than 15 s. The strength of the responses indicates bubble radii of $\approx$2.5 cm and it is extremely unlikely that they result from free gas bubbles. We are thus led to the conclusion that the many and varied *in situ* ping responses we have observed are the result of the high-order resonant response of fish swimbladders. With further theoretical and experimental study, the effects of high-order resonances could be a useful aid in estimating fish size and possibly species. The presence of such resonances, predicted theoretically for real swimbladder shapes, has been reported recently by Francis and Foote (2003).

These results on high-order resonance effects may also have relevance to fisheries biomass estimation. This relies on the integral of the acoustic response of the target under investigation and also uses target strength to convert such integrals into biomass. Near high-order resonances the integral can be markedly affected by the skirt response of the target (Figs. 3 and 4), particularly those with very slow decay times (Fig. 4). However, our results also suggest that this effect will only be significant for fishes smaller than $\approx$15–20 cm.

In the case of target strength, it is usual to assume a simple logarithmic relationship between target strength (TS) and fish length (L), namely,

$$\mathrm{TS} = A \log_{10}(L) + B. \tag{3}$$

With this formulation, target strength increases monotonically with fish length. The total biomass is then computed from the integral of all the acoustic ping responses and scaled using fish length and target strength data. It is clear from Fig. 3 that the integral response of a fish with a bladder of radius 88% would be significantly larger than that of a fish with a bladder of radius 75% or 105%. It is also noteworthy that the response for a radius of 75% is greater than that for 105%.

ACKNOWLEDGMENT

We would like to thank John McKoy for allocating funding for this research under Contract No. FSF2000A, task 14.

Anderson, V. C. (**1950**). ''Sound scattering from a fluid sphere,'' J. Acoust. Soc. Am. **22**, 426–431.

Barr, R., and Coombs, R. F. (to be published). ''Target rate of change of phase: A new dimension for fish and plankton target identification,'' J. Acoust. Soc. Am.

Barr, R., Coombs, R., Doonan, I., and McMillan, P. (**2002**). ''Target identification of oreos and associated species,'' Final Research Report for Ministry of Fisheries project OEO2000/01B, Objective 1 (unpublished report held by the Ministry of Fisheries, Wellington).

Batzler, W. E., and Pickwell, G. V. (**1970**). ''Resonant acoustic scattering from gas-bladder fishes,'' in *Proceedings of an International Symposium on Biological Sound Scattering in the Ocean*, edited by G. B. Farquhar (U.S. Govt. Printing Office, Washington, DC, 1970), pp. 168–179.

Coombs, R. F., Macaulay, G. J., Knol, W., and Porritt, G. (**2003**). ''Configurations and calibrations of 38 kHz fishery acoustic survey systems, 1991–2000,'' New Zealand Fisheries Assessment Report 2003/49.

Feuillade, C., and Werby, M. F. (**1994**). ''Resonances of deformed gas bubbles in liquids,'' J. Acoust. Soc. Am. **96**, 3684–3692.

Francis, D. T. I., and Foote, K. G. (**2003**). ''Depth-dependent target strengths of gadoids by the boundary-element method,'' J. Acoust. Soc. Am. **114**, 3136–3146.

Furusawa, M. (**1988**). ''Prolate spheroidal models for predicting general trends of fish target strength,'' J. Acoust. Soc. Jpn. (E) **9**, 13–23.

Johnson, R. K. (**1977**). ''Sound scattering from a fluid sphere revisited,'' J. Acoust. Soc. Am. **61**, 375–377.

Kloser, R. J., Ryan, T. E., Williams, A., and Soule, M. (**2000**), ''Development and implementation of an acoustic survey of orange roughy in the Chatham Rise spawning box from a commercial factory trawler, *FV Amaltal Explorer*,'' Report 597.64, Commonwealth Scientific and Industrial Research Organisation, Marine Research Division, Hobart, Tasmania.

Love, R. H. (**1978**). ''Resonant acoustic scattering by swimbladder-bearing fish,'' J. Acoust. Soc. Am. **64**, 571–580.

Macaulay, G. J. (**1997**). ''Estimation of *in-situ* acoustic target strength of hoki,'' Final Research Report for Ministry of Fisheries project POAC04 (unpublished report held by the Ministry of Fisheries, Wellington).

MacLennan, D. N. (**1981**). ''The theory of solid spheres as sonar calibration targets,'' Scottish Fisheries Research Report, No. 22, ISSN 0308-8022.

Marshall, N. B. (**1960**). ''Swimbladder structure of deep-sea fishes in relation to their systematics and biology,'' *Discovery Reports*, Vol. XXXI (Cambridge U.P., Cambridge).

McCartney, B. S., and Stubbs, A. R. (**1970**). ''Measurements of the target strength of fish in dorsal aspect, including swimbladder resonance,'' in *Proceedings of an International Symposium on Biological Sound Scattering in the Ocean*, edited by G. B. Farquhar (U.S. Govt. Printing Office, Washington, DC, 1970), pp. 180–211.

McClatchie, S., and Coombs, R. F. (**2000**). ''*In-situ* target strength measurements of orange roughy (*Hoplostethus atlanticus*),'' Final Research Report

for Ministry of Fisheries project ORH9801, Objective 1 (unpublished report held by the Ministry of Fisheries, Wellington).
Medwin, H., and Clay, S. C. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, London), pp. 276–286.
Rayleigh, Lord (J. W. Strutt) (**1896**). *The Theory of Sound* (Dover, New York, 1945), Vol. 2, pp. 202–289.
Rudgers, A. J. (**1967**). ''Techniques for numerically evaluating the formulas describing monostatic reflection of acoustic waves by elastic spheres,'' NRL Report 6551, Naval Research Laboratory, Washington, DC.
Sand, O., and Hawkins, A. D. (**1973**). ''Acoustic properties of the cod swimbladder,'' J. Exp. Biol. **58**, 797–820.
Silbiger, A. (**1963**). ''Scattering of sound by an elastic prolate spheroid,'' J. Acoust. Soc. Am. **35**, 564–570.
Stansfield, D. (**1991**). *Underwater Electroacoustic Transducers* (Bath U.P., Bath), Chap. 4, pp. 73–103.
Stanton, T. K., and Chu, D. (**1998**). ''Sound scattering by several zooplankton groups, II. Scattering models,'' J. Acoust. Soc. Am. **103**, 236–253.
Warren, J. D., Stanton, T. K., Benfield, M. C., Wiebe, P. H., Chu, D., and Sutor, M. (**2001**). ''*In situ* measurements of acoustic target strengths of gas-bearing siphonophores,'' ICES J. Mar. Sci. **58**, 740–749.
Yeh, C. (**1967**). ''Scattering of acoustic waves by a penetrable prolate spheroid. I. Liquid prolate spheroid,'' J. Acoust. Soc. Am. **42**, 518–521.

Identification of laser generated acoustic waves in the two-dimensional transient response of cylinders

Y. Pan[a)]
Institute of Acoustics, Tongji University, 200092, Shanghai, People's Republic of China

C. Rossignol and B. Audoin
Laboratoire de Mécanique Physique, UMR CNRS 8469, Université Bordeaux 1, 33405 Talence, France

(Received 26 January 2005; revised 1 April 2005; accepted 4 April 2005)

The published model [Appl. Phys. Lett. **82**, 4379–4381 (2003)] for the two-dimensional transient wave propagation in a cylinder is modified to avoid the inherited integration of the numerical inverse scheme. The Fourier series expansion is introduced for one spatial coordinate to resolve the transient response problem: theoretical radial displacements in either the ablation or the thermoelastic regime are obtained with little numerical noise and short computation time. The normal mode expansion method fails to deliver results with the same accuracy. Acoustic waves are fully identified by the ray trajectory analysis. These identified waves are further verified on the experimental results observed with the laser ultrasonic technique. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1921527]

PACS numbers: 43.35.Cg, 43.20.Px, 43.20.Bi [YHB] Pages: 3600–3608

I. INTRODUCTION

A cylinder with its unique geometry is a basic target for many research fields. Specifically, understanding the elastic wave propagation is very important for the nondestructive evaluation community, since fibers, composite rods, and rotating axes of a machine have the cylindrical structure.

Surface acoustic or Rayleigh wave propagating on an isotropic and homogeneous cylinder was considered in 1927 by Sezawa,[1] who obtained its characteristic equation and calculated the dependence of its velocity on the radius of the cylinder. More detailed studies were reported in 1967 by Viktorov,[2] who calculated the dispersion curve and made comparison to the experimental data observed with conventional piezoelectric transducers. Higher Rayleigh-type waves were termed as ''Whispering-gallery modes'' by Uberall in 1973.[3] Due to the coupling difficulty of such transducers, few experimental reports on the wave propagation for a curved medium had been published until the development of the laser ultrasonic technique,[4] in which ultrasonic waves are both generated and detected with lasers. With the remarkable features of noncontact, high spatial and temporal resolutions introduced by this technique, various studies on materials of curved surfaces have been carried out. For instance, Rayleigh wave propagating on a sphere was observed experimentally in 1988,[5] and a further detailed study has been reported later.[6] Rayleigh wave propagating on a cylinder was also studied by this technique.[7,8] The finite element method has been used to predict the bulk and surface wave propagations when the laser beam was focused by a cylindrical lens.[9] Theoretical solutions have also been reported for the two- and three-dimensional (2D and 3D) transient guided wave propagation in a hollow cylinder.[10,11] The velocities of the Rayleigh wave propagating on steel rods receiving different heat treatments were measured to characterize the surface residual stress of the cylindrical parts.[12] The surface wave propagation on a cylinder has also been studied to detect surface break defects.[13]

Recently, the authors have presented a model[14] to solve the transient elastic wave propagation generated by a laser line pulse in a cylinder. Under either the ablation or the thermoelastic regime, the calculated radial displacements are in good agreement with that observed with the laser ultrasonic technique for aluminum cylinders. With the theoretical wave forms, the surface wave and several bulk waves such as the direct or reflected longitudinal waves were well identified. However many acoustic waves with obvious relative large amplitude were still not understood, and these waves may be very helpful for the inverse problem when the known waves fail to deliver enough information for NDE purposes such as the application to elastic constants measurement.[15] Even though no application is considered, to reveal the intrinsic rich information in the dynamical response is still important, since it will deepen the current understanding about the wave generation and propagation in cylindrical structures.

However, the published model necessitated an integration along one wave number for the numerical inverse scheme, thus it required a quite amount of calculation time on a standard computer, and it also induced numerical noise (see Figs. 1–3 in Ref. 14). Moreover, this inverse scheme became worse when dealing with the similar problem under a laser point source,[16] because the integration became 2D for the 3D problem. Although the normal mode expansion, also known as the eigenfunction expansion, which avoids the integration, has successfully been applied to solve the 3D wave propagation under the laser generation in a hollow cylinder,[11] it is still unclear whether it can be applied to cylindrical rods.

In this paper, our previously published model predicting the acoustic wave generated in a cylinder by a laser line source is improved to avoid the inherited integration of the Fourier transform. Transformed solutions of the modified model with a novel numerical inverse scheme are presented.

a)Electronic mail: ypan@mail.tongji.edu.cn

 0001-4966/2005/117(6)/3600/9/$22.50

Better numerical results are obtained allowing a detailed wave analysis. A large amount of previously unknown waves are compared and identified on various numerical results including the normal mode expansion solution. The identified waves are further verified by the experimental results observed with the laser ultrasonic technique.

II. GENERAL FORMULATION

Let us consider a homogeneous and isotropic cylinder of infinite length, radius a, and density ρ. The symmetrical axis of the cylinder is assumed to coincide with the z axis of its cylindrical coordinates (r,θ,z). A pulsed laser is used to generate acoustic waves in this material. If a cylindrical lens is used to focus the laser beam, this laser can be modeled as a line-like acoustic source, which lies at the boundary of the cylinder, $r=a$, and shows infinite extension along the z direction. To denote the detection position on the cylinder, cylindrical coordinates (a,θ) are chosen, considering $\theta=0°$ for the source position. Here, an experimental technique, generally an optical detection technique,[4] is applied to measure these acoustic waves.

Let λ and μ denote the two Lamé coefficients for the isotropic cylinder. Owing to the symmetry imposed by the source shape, this problem shows invariance along the z direction. The nonzero components of the displacement vector depend on two spatial variables r, θ and on time t. These components, denoted u_r and u_θ, can be expressed in terms of a scalar potential φ and the nonzero component of a vector potential $\mathbf{\Psi}=(0,0,\psi)$ as[17]

$$u_r(r,\theta,t)=\frac{\partial\varphi}{\partial r}+\frac{1}{r}\frac{\partial\psi}{\partial\theta},$$
$$u_\theta(r,\theta,t)=\frac{1}{r}\frac{\partial\varphi}{\partial\theta}-\frac{\partial\psi}{\partial r}, \tag{1}$$

where φ and ψ are governed by the following elastic wave motion equations:

$$\frac{\partial^2\varphi}{\partial r^2}+\frac{1}{r}\frac{\partial\varphi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\varphi}{\partial\theta^2}=\frac{\rho}{\lambda+2\mu}\frac{\partial^2\varphi}{\partial t^2},$$
$$\frac{\partial^2\psi}{\partial r^2}+\frac{1}{r}\frac{\partial\psi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2}=\frac{\rho}{\mu}\frac{\partial^2\psi}{\partial t^2}. \tag{2}$$

Then, two components σ_{rr} and $\sigma_{r\theta}$ of the stress tensor at any point of the surface must comply with the following boundary conditions:

$$\sigma_{rr}|_{r=a}=-F_0\delta(t)\delta(\theta),$$
$$\sigma_{r\theta}|_{r=a}=0, \tag{3}$$

for the ablation generation,[18] or

$$\sigma_{rr}|_{r=a}=0,$$
$$\sigma_{r\theta}|_{r=a}=-F_0h(t)\delta'(\theta), \tag{4}$$

for the thermoelastic generation.[19] F_0 is a certain loading in $\mathrm{N}\,\mu\mathrm{s}\,\mathrm{m}^{-2}$ for Eq. (3) and in $\mathrm{N}\,\mathrm{m}^{-2}$ for Eq. (4), respectively. It is related to the intensity of the laser line pulse, and its magnitude is fixed to unity for the corresponding calculations in the following part. Here a delta function of time $\delta(t)$ and a Heaviside step function of time $h(t)$ are used for the ablation and thermoelastic generations, respectively; and $\delta'(\theta)$ denotes the derivative of the delta function $\delta(\theta)$. In Eq. (3), a delta force is postulated in time and space to represent sudden normal loading in the ablation regime. A Heaviside step function in time is considered in the thermoelastic regime [Eq. (4)] since thermal diffusion is neglected.[19] Owing to the interface, a dipolar force is considered in Eq. (4) for the source shape.[20] The number of clockwise or counterclockwise roundtrips of the generated waves as considered in the previous model is dismissed by setting the range for θ from $-\pi$ to π instead of $-\infty$ to ∞. This modification accounts for the periodicity of the cylindrical structure as presented in the following section.

III. TRANSFORMED SOLUTIONS AND NUMERICAL INVERSE SCHEME

To avoid the integration along the wave number ν ($\nu=k_\theta a$, where k_θ is the component of the wave vector $\mathbf{k}$ along the θ direction) necessitated by the numerical inverse scheme in Ref. 14, the Fourier series expansion is introduced for θ accounting for the periodicity of the geometry. The Fourier transform is considered for time, and the angular frequency ω is thus introduced. Doing so, the actual displacement field $u_i(r,\theta,t)$, $i=r$ or θ, is now transformed to $U_i(r,\nu,\omega)$, and it can be then expressed as the following inverse form:

$$u_i(r,\theta,t)=(2\pi)^{-2}\int_{-\infty}^{+\infty}\left(\sum_{\nu=-\infty}^{+\infty}U_i(r,\nu,\omega)e^{-j\nu\theta}\right)e^{j\omega t}\,d\omega. \tag{5}$$

Applying this modified transform, the wave motion equations and the boundary equations are linearized, providing explicit forms for the potentials φ and ψ under either the ablation or the thermoelastic regime. Simple explicit forms of the displacement components at the cylindrical surface are then obtained for the ablation regime

$$U_r(a,\nu,\omega)=-\frac{F_0a}{2\mu D(\omega,\nu)}\left\{\nu^2-\left(\frac{k_T^2a^2}{2}+B_T\right)B_L\right\},$$
$$U_\theta(a,\nu,\omega)=\frac{j\nu F_0a}{2\mu D(\omega,\nu)}\left(\nu^2-\frac{k_T^2a^2}{2}-B_LB_T\right), \tag{6}$$

and for the thermoelastic regime

$$U_r(a,\nu,\omega)=-\frac{\nu^2F_0aH(w)}{2\mu D(\omega,\nu)}\left(\nu^2-\frac{k_T^2a^2}{2}-B_LB_T\right),$$
$$U_\theta(a,\nu,\omega)=\frac{j\nu F_0aH(w)}{2\mu D(\omega,\nu)}\left\{\nu^2-\left(\frac{k_T^2a^2}{2}+B_L\right)B_T\right\}, \tag{7}$$

where

$$D(\omega,\nu)=(\nu^2-k_T^2a^2/2)^2-\nu^2+k_T^2a^2(B_L+B_T)/2+(1-\nu^2)B_LB_T,$$
$$B_L=k_La\,J_\nu'(k_La)/J_\nu(k_La),$$
$$B_T=k_Ta\,J_\nu'(k_Ta)/J_\nu(k_Ta). \tag{8}$$

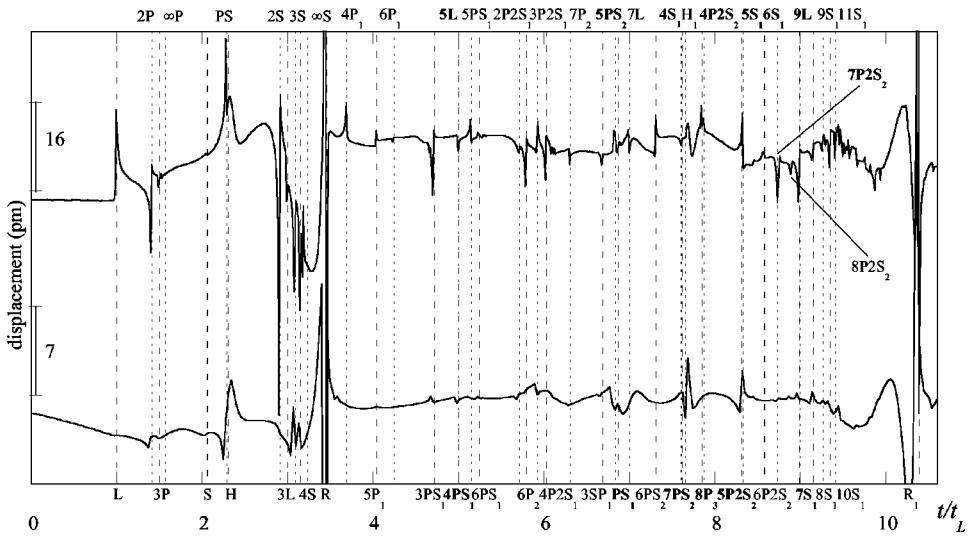


FIG. 1. Acoustic waves obtained by the Fourier series expansion for aluminum cylinders at the epicenter of the laser line source generation under either the ablation (top) or the thermelastic (below) regime. Symbols at the top or bottom of each dashed line, respectively, denote the wave marked by the dense or sparse dashed line.

In Eqs. (6)–(8), $k_L=\omega\sqrt{\rho/(\lambda+2\mu)}$ and $k_T=\omega\sqrt{\rho/\mu}$ are the wave numbers of the longitudinal and transverse waves, respectively. Additionally, $H(\omega)$ is the Fourier transform of Heaviside step function $h(t)$, and $J'_\nu(x)$ is the derivative of the Bessel function $J_\nu(x)$.

Now, let us focus on the calculation of the discrete summation in the inverse transform. For an elastic material without any attenuation, this summation may meet discontinuities for particular ν, i.e., $k_\theta a$, values. They correspond to poles associated with the zeros of the dispersion equation

$$D(\omega,\nu)=0, \tag{9}$$

which describe the cylindrical Rayleigh waves[2] and Whispering Gallery waves.[3] Thus the inverse transform cannot be carried out numerically. Therefore, complex frequencies are considered introducing a small constant imaginary part δ as presented in a previous paper.[14] Doing so, ω becomes complex, and the poles are moved away from the real axis of the variable ν. To distinguish from the previous method of the integral Fourier transform, the current method is called the Fourier series expansion in the following. Various numerical results are going to demonstrate its benefit in Sec. IV.

IV. NUMERICAL RESULTS AND WAVE IDENTIFICATION

The calculated radial displacements generated by the laser line pulse under either the ablation or the thermoelastic regime are shown in Fig. 1. The two displacements are obtained by the presented method, the Fourier series expansion method, for aluminum cylinders with velocities of longitudinal wave $V_L=6.40$ km/s, and shear wave $V_T=3.11$ km/s, at an observation angle θ of 180°. The value $\delta=0.06$ rad μs^{-1} has been chosen for the numerical inversions. Through this paper, the time scale is made dimensionless by dividing the time t by a factor t_L, the transit time of the direct longitudinal wave for this epicenter position ($t_L=2a/V_L$). This treatment favors the further comparisons, since the time scale remains unchanged for the same material with a different radius. Moreover, alternate sparse and dense dashed lines are repeatedly displayed to indicate the time arrivals of various waves. The arrival time of each wave is calculated accurately based on its corresponding ray trajectory shown in Fig. 2. Symbols at the top or bottom of each pair of dashed lines, respectively, represent the name of the wave marked by the dense or sparse dashed line. This representation holds for the similar figures through this paper.

First let us analyze the acoustic waves generated under the ablation regime. The cylindrical Rayleigh wave R and its one-roundtrip sequence R_1 are clearly visible with a large amplitude. This subscript 1 expresses that the wave path has turned around the cylindrical center for one cycle besides a half circle from the generation to the detection for the chosen observation angle, and in the following text it indicates the same sense by subscribing the name of any wave. To better visualize other waves with small amplitudes, the two Rayleigh waves are not fully displayed here. But it is still clear that R appears as a bipolar, that R_1 spreads in two successive periods, and that the components of the low frequency part travel relatively fast. The evolution of the shape of this Rayleigh wave during its propagation can be explained by its dispersive nautre.[2,21] Moreover, its group wave velocity cannot be precisely determined because of its dispersiveness. Thus its transit time is estimated using Rayleigh velocity at the boundary of a halfspace just for the purpose of marking the time arrival of the identified wave.

The direct longitudinal wave L and its diametrically reflected sequences [$3L$, $5L$, $7L$, and $9L$ in Fig. 2(a)] can be easily identified in Fig. 1, even though the sequences $5L$ and $9L$ may be overlapped with a converted wave $4PS_1$ [a wave resulting of the mode conversion between longitudinal and transverse waves, see Fig. 2(f)] and a circumferentially reflected transverse wave $7S_1$ [Fig. 2(c)] respectively. Waves $4PS_1$ and $7S_1$ will be analyzed in the following part. All the overlaped waves in Fig. 1 are highlighted in bold characters. The longitudinal wave arrivals are 1, 3, 5, 7, and 9 in the normalized time scale, and their ray paths are shown as nL ($n=1,3,...,9$) in Fig. 2(a). Meanwhile, the direct transverse wave T with the same ray path as L also shown in Fig. 2(a) is hardly visible in Fig. 1. This difference can be explained by the different directivities[4,22] of the longitudinal and transverse wave under this ablation generation regime. As shown in Figs. 3(a) and 3(b), the directivity of the longi-

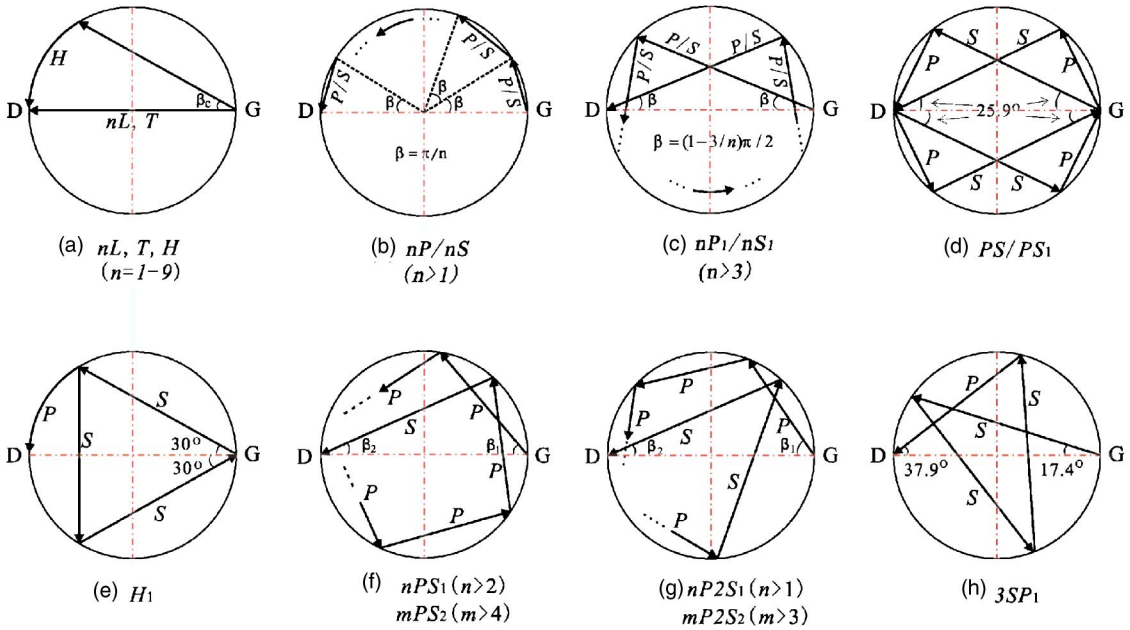


FIG. 2. Ray trajectories of longitudinal (nL,nP,nP_1), transverse (T,nS,nS_1), head (H,H_1), and converted (PS,PS_1,nPS_1,mPS_2,$nP2S_1$,$mP2S_2$,$3SP_1$) waves. Here the subscript 1 or 2 indicates that the path has turned around the cylinder center for one or two cycles, respectively, in addition to a half cycle for the 180° observation angle. β_c is the critical angle for the transverse wave path. β_1 and β_2 are angles related to the longitudinal and transverse wave path, respectively. P denotes the longitudinal path, and S symbolizes the transverse or shear wave path.

tudinal wave has a maximum at 0°, the epicenter direction angle, while that of the transverse wave has a minimum (zero amplitude) at the same angle. It is interesting to notice that L is a monopolar with a positive polarization, whereas $3L$ and $7L$ are monopolars with negative polarizations in Fig. 1. Since there is a loss of energy between each reflection, their corresponding amplitudes are in decreasing orders as appears in the figure, and this phenomenon remains for the following waves in the similar sequences.

The circumferentially reflected longitudinal waves [$2P$, $3P$ in Fig. 2(b)], their one-roundtrip sequences [$4P_1$, $5P_1$, $6P_1$ in Fig. 2(c)], their two-roundtrip sequences ($6P_2$,$7P_2$) are all visible in Fig. 1. And one of their three-roundtrip sequences $8P_3$ is also visible in Fig. 1, even though it is very close to a converted wave $4P2S_2$ [Fig. 2(g)]. Following the subscript 1 defined previously, the subscripts 2 and 3 denote that the wave path has turned around the cylindrical center for two and three cycles, respectively, besides the additional half cycle for the chosen observation angle. Waves $2P$ and $3P$ have the ray trajectory denoted as nP ($n>1$) in Fig. 2(b) with the corresponding angle $\beta=\pi/n$. Waves $4P_1$, $5P_1$, and $6P_1$ have the ray trajectory noted as nP_1 ($n>3$) in Fig. 2(c) with the corresponding angle $\beta=(1-3/n)\pi/2$. Here, the ray path has turned for one more cycle as indicated by the subcript 1. Similarly, $6P_2$, $7P_2$, and $8P_3$ have their paths turning around the center for two and three more cycles, respectively, and their ray trajectories (not displayed in Fig. 2) are quite similar to that of nP_1 turning around the center with more cycles. High order waves such as $4P$, $7P_1$, $8P_2$, and $9P_3$ are hardly visible in Fig. 1, because their amplitude is weaker compared to their lower order counterparts, since there is an energy loss between each reflection. The normalized values of the time arrivals are $n\sin(\pi/2n)$, $n\cos[(1-3/n)\pi/2]$, $n\cos[(1-5/n)\pi/2]$, and $n\cos[(1-7/n)\pi/2]$ for

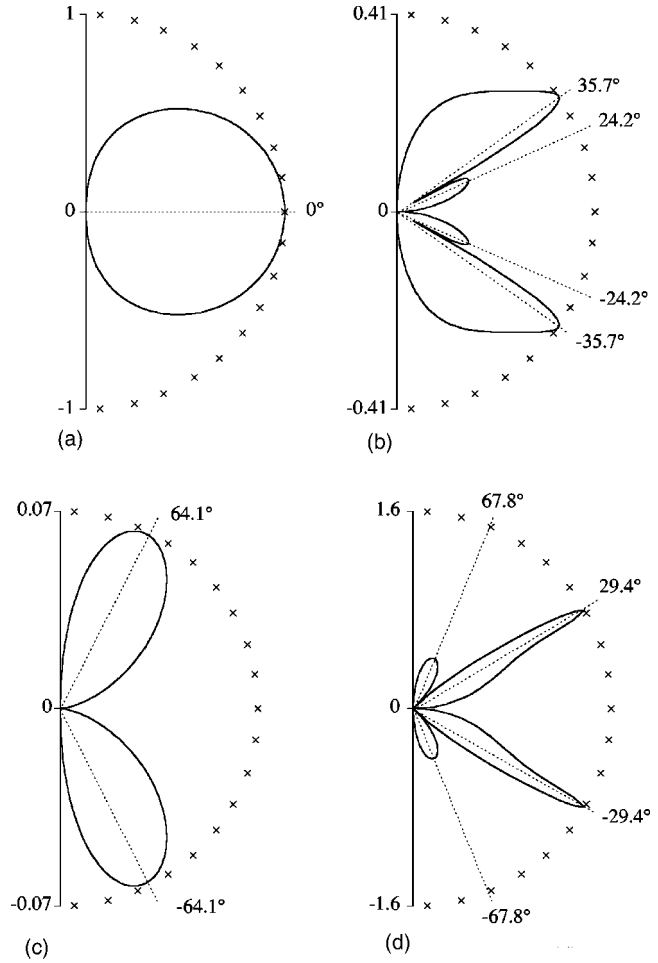


FIG. 3. Directivity or angular dependence of (a) longitudinal, (b) transverse waves under the ablation generation and (c) longitudinal, (d) transverse waves under the thermoelastic generation.

the waves $nP\,(n>1)$, $nP_1\,(n>3)$, $nP_2\,(n>5)$, and $nP_3\,(n>7)$, respectively. When $n\to\infty$, these four values turn out to be $\pi/2$, $3\pi/2$, $5\pi/2$, and $7\pi/2$, correspondingly. Obviously, ∞P, ∞P_1, ∞P_2, and ∞P_3 are the longitudinal waves propagating along the circumference as lateral or skimming waves.[17] Unfortunately, they are not visible as indicated by the ∞P in Fig. 1, because the directivity of the longitudinal wave in Fig. 3(a) indicates that no generation occurs at the radiation angle of 90°.

Similarly, the circumferentially reflected transverse waves [$2S$, $3S$, and $4S$ in Fig. 2(b)] and their one-roundtrip sequences [$5S_1,\ldots,11S_1$ in Fig. 2(c)] are clearly visible in Fig. 1, despite the overlaps between $5S_1$ and $5P2S_2$, $6S_1$ and $7P2S_2$. Waves $5P2S_2$ and $7P2S_2$ shown in Fig. 2(g) will be analyzed later. A one-roundtrip sequence $4S_1$ is hard to identify, because it has a small relative amplitude of 0.16 at the angle of radiation 22.5° in the directivity diagram of Fig. 3(b) (0.41 for $5S_1$ at 36°), and it is very close to a converted wave $7PS_2$ in Fig. 2(f). As shown in Figs. 2(b) and 2(c), $nS\,(n>1)$ and $nS_1\,(n>3)$ have the same ray paths as their longitudinal counterparts, and their time arrivals are easily obtained by multiplying that of their longitudinal counterparts with a factor V_L/V_T, the ratio of the longitudinal and transverse wave velocities. High order waves such as $5S$ and $12S_1$ are hard to distinguish, because they suffer the same decrease in amplitude as the longitudinal counterparts. The time delay between these sequential waves shortens so much that they cannot be separated further. When $n\to\infty$, ∞S and ∞S_1 are skimming transverse waves propagating along the surface as their longitudinal counterparts. Again, they are not visible as indicated by the ∞S in Fig. 1, because the directivity of the transverse wave also shows no generation at the radiation angle of 90° in Fig. 3(b).

Unlike the diametrically reflected longitudinal wave, the head wave H with the ray path shown in Fig. 2(a), and its one-roundtrip sequence H_1 with the ray path shown in Fig. 2(e), are visible in Fig. 1 with a broad shape. H is a positive monopolar, and H_1 is a bipolar. For the H wave, the dense dashed line indicates the arrival of the longitudinal wave converted from the transverse wave propagating at the critical angle β_c. Later contributions after this line are the conversions from the transverse waves propagating at an angle greater than β_c. Here β_c equals 29.1° for the chosen aluminum cylinder. This longitudinal wave propagates along the cylindrical surface as a skimming wave. The maximum amplitude of this H wave in Fig. 1 corresponds to the maximum amplitude in the directivity of the transverse wave with radiation angles ±35.7° in Fig. 3(b). This directivity also has a side lobe centered at ±24.2°, which may explain the started arrival of H_1 in Fig. 1, since the contribution of the main lobe is behind the marked sparse dashed line.

Close to the wave H, a conversion wave PS can be identified as a monopolar pulse with a large amplitude. This wave was not found in our previous numerical wave form (see Fig. 1 in Ref. 14), as the numerical noise hindered the identification. It has the ray path shown in Fig. 2(d). It is mainly attributed by the wave conversion from the longitudinal P to transverse S waves, as the directivity of P wave shows a relative amplitude of 0.51 in Fig. 3(a) versus 0.16 in Fig. 3(b) for the directivity of S wave at their corresponding radiation angles of ±64.1° and ±25.9°. Naturally, its one-roundtrip sequence PS_1 with the ray path also shown in Fig. 2(c) can also be identified, despite the overlapping with another converted wave $5PS_2$.

Other conversion waves with more complicated ray paths are also identified. They are $nPS_1\,(n=3-6)$ and $mPS_2\,(m=5,6,7)$ waves with one path of the shear wave, $nP2S_1\,(n=2,3,4)$ and $mP2S_2\,(m=4-8)$ waves with two paths of the shear wave, and $3SP_1$ wave with three paths of the shear wave. Here $7P2S_2$ is the pulse overlapped with $6S_1$, and $8P2S_2$ is the peak between $6S_1$ and $7S_1$. Their corresponding ray trajectories are found in Figs. 2(f)–2(h). The displayed ray path is only a specific case, because the time arrival remains unchanged with the different order of the combination while maintaining the number for P and S as indicated. In Figs. 2(f) and 2(g), β_1 and β_2 are the angles related to the longitudinal and transverse wave path, respectively. For the combination of one shear wave path, β_1 and β_2 are 25.9° and 12.3° for $3PS_1$, 40.4° and 18.4° for $4PS_1$, 49.7° and 21.7° for $5PS_1$, 56.0° and 23.8° for $6PS_1$, 16.4° and 7.9° for $5PS_2$, 27.8° and 13.1° for $6PS_2$, 36.2° and 16.7° for $7PS_2$, respectively. For the combination of two shear wave paths, β_1 and β_2 are 30.7° and 14.3° for $2P2S_1$, 46.3° and 20.6° for $3P2S_1$, 55.7° and 23.7° for $4P2S_1$, 18.2° and 8.7° for $4P2S_2$, 30.3° and 14.2° for $5P2S_2$, 39.1° and 27.8° for $6P2S_2$, 45.6° and 20.3° for $7P2S_2$, 50.7° and 22.1° for $8P2S_2$, respectively. For $3SP_1$, β_1 is 37.9°, and β_2 is 17.4°. These β_1 values are within the wide range of −86.7° to 86.7° accounting for the P wave directivity, those for β_2 are also within the range of −87.3° to −4.9° and 4.9° to 87.3° for the S wave directivity. These limits are the radiation angle for which the amplitude is greater than one-tenth of the maximum amplitude. High order waves such as $7PS_1$, $8PS_2$, $5P2S_1$, $9P2S_2$, and $4SP_1$ are hard to distinguish in Fig. 1, owing to the same reason as for the circumferentially reflected longitudinal and transverse waves.

Now most of the elastic waves in the wave form are identified, but some of them may be totally mistaken as other waves, owing to overlapping, especially for waves with different subscript. First let us verify the subscript associated to each wave. Because the current model cannot predict the waves with the different subscript separately, our previously published model is adopted in spite of its weakness. This model is able to predict waves with their trajectories turning around the cylindrical center at a chosen number of cycles. As shown in Fig. 4, waves with trajectories up to two (top) or one (below) clockwise and counterclockwise roundtrips are obtained for aluminum cylinders at the epicenter of the laser line source generation under the ablation regime. All the waves highlighted in bold characters with subscript 2 [$6P_2$,$7P_2$,$mPS_2\,(m=5,6,7)$,$mP2S_2\,(m=4-8)$] are visible in the top wave form, while they are invisible in the low wave form. Another highlighted wave $8P_3$ with the subscript 3, visible in Fig. 1, is no longer visible in both wave forms in Fig. 4. Moreover, further calculations (not presented) also show that all waves with the subscript 1–3 in Fig. 1 disappear when the number of cycles is chosen to be zero. Therefore, we can conclude that each identified wave has its right

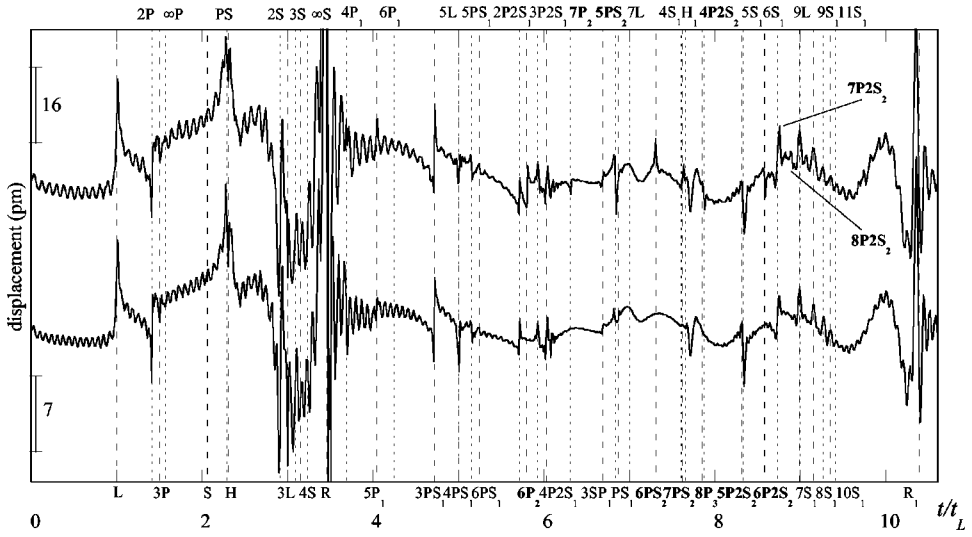


FIG. 4. Acoustic waves with trajectories up to two (top) or one (below) clockwise and counterclockwise roundtrips in the radial displacements obtained by the integration along ν for aluminum cylinders at the epicenter of the laser line source generation under the ablation regime. Symbols at the top or bottom of each dashed line, respectively, denote the wave marked by the dense or sparse dashed line.

subscript. Although this method is successful in separating waves with different subscript, the novel method is far better in view of accuracy and efficiency. Numerical noises as appearing in Fig. 4 are completely invisible in Fig. 1. The accuracy of the computation is greatly improved by this new method. Moreover, it is more than 20 times faster than the previous method.

Now let us verify the overlaps in Fig. 1 such as $5L$ with $4PS_1$ and $9L$ with $7S_1$. Similar calculations were carried for a copper cylinder with velocities of longitudinal wave $V_L = 4.70\,\mathrm{km/s}$, and shear wave $V_T = 2.26\,\mathrm{km/s}$. Other parameters for the calculations remain unchanged for this verification. As shown in the top wave form in Fig. 5 with bold characters, $5L$ does not overlap with $4PS_1$, and $9L$ is clearly separated with $7S_1$. Moreover $4S_1$ is also apart from $7PS_2$, which clearly indicates that $4S_1$ is hardly visible. And $6S_1$ is slightly apart from $7P2S_2$, which is a small peak just before $6S_1$. In addition, the wave $8P2S_2$ is the peak just after $6S_1$ and before $9L$. All other waves identified for aluminum cylinders can be found in this figure. Obviously, the method of wave identification for aluminum cylinders is also valid for the copper cylinders, and this emphasizes the validness of the identification method. Following a similar procedure, the visibility of other possible waves such as $2PS$, $3S2P_1$, $7PS_3$, and $10P_4$ are also verified. Other possible arrivals of one wave or of a supposition of several waves could also be identified in the left unstated part of the wave forms in Figs. 1, 4, and 5, even though it is not a trivial matter.

Now let us analyze the waves generated under the thermoelastic regime. As shown in Fig. 3(c), for this regime, the directivity diagram of the longitudinal wave is two symmetrical lobes maximizing at $\pm 64.1°$ instead of one lobe maximizing at 0° for the ablation regime. It has a narrow range from $-88.9°$ to $-14.3°$ and from 14.3° to 88.9° instead of $-86.7°$ to 86.7°. There is no radiation of longitudinal wave at the epicenter direction, thus the direct longitudinal wave L and its diametrically reflected sequences $3L$, $5L$, $7L$, and $9L$ are not visible in Fig. 1. Although waves generated by this radiation such as nP, nP_1, nP_2, nPS_1, nPS_2, $nP2S_1$, $3SP_1$ still exist, their amplitudes are less than their counterparts for the ablation regime, due to the small amplitude in the directivity (maximum value is 0.07 at $\pm 64.1°$). As shown in Fig. 3(d), the directivity of the shear wave is also different from that of the ablation regime. It has two symmetrical lobes from $-42.1°$ to $-4.6°$ and from 4.6° to

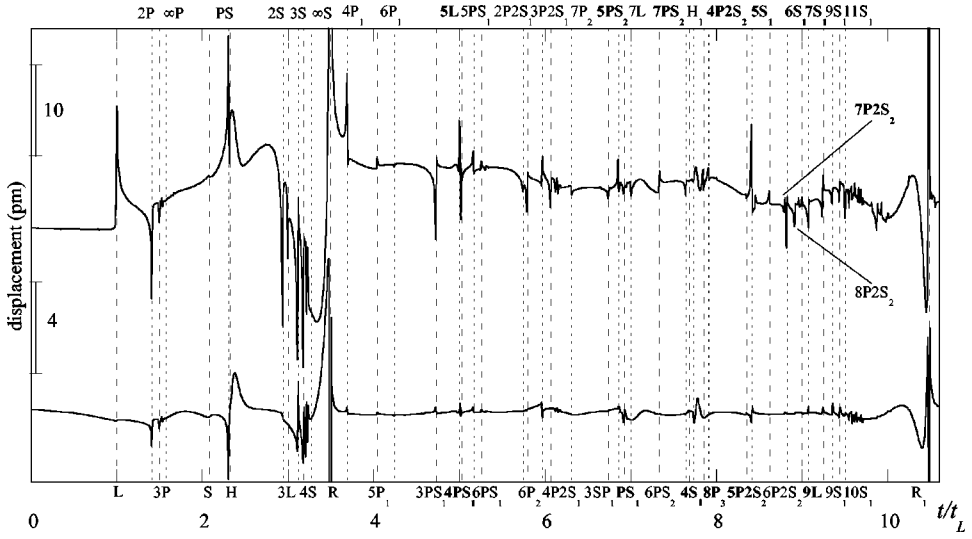


FIG. 5. Acoustic waves obtained by the Fourier series expansion for copper cylinders at the epicenter of the laser line source generation under either the ablation (top) or the thermelastic (below) regime. Symbols at the top or bottom of each dashed line, respectively, denote the wave marked by the dense or sparse dashed line.

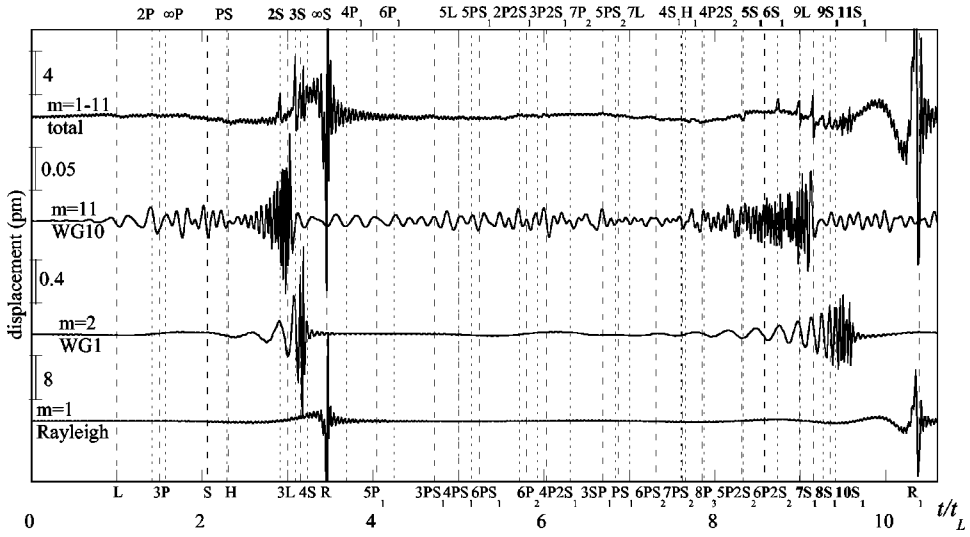


FIG. 6. Acoustic waves obtained by the normal mode expansion for aluminum cylinders at the epicenter of the laser line source generation under the ablation regime with the summation of 11 modes. Symbols at the top or bottom of each dashed line, respectively, denote the wave marked by the dense or sparse dashed line.

42.1°, maximizing at ±29.4° instead of the corresponding ranges from −87.3° to −29.1° and from 29.1° to 87.3°, maximizing at ±35.7° for the ablation regime. It is narrower with higher amplitudes. The two side lobes maximize at ±67.8° with the relative amplitude of 1.61 versus 0.41 for the ablation at ±24.2°. This explains the increase in the wave H_1, and the hold of the wave $5S_1$, while other waves decrease in Fig. 1. Overall, these differences between the two regimes emphasize the validness of the wave identification by the ray trajectory analysis. As shown in Fig. 5, this phenomenon also happens for the copper cylinder, further emphasizing the validation of the method for the wave calculation and identification.

V. COMPARISON WITH THE NORMAL MODE EXPANSION

Since the normal mode expansion method is successful in predicting the 2D and 3D transient guided wave propagations in a hollow cylinder, let us verify whether it can be applied to a simple case for the cylindrical rod.

The three major steps reported by Liu and Qu[10] have been followed to obtain the 2D transient response under the laser ablation generation. First, the corresponding steady-state problem was solved to obtain the individual steady-state vibration mode $\vec{H}^{(i)}_{m\nu}(r,\theta)$,

$$\begin{cases} \vec{U}^{(1)}_{m\nu}(r,\theta) \\ \quad = \begin{Bmatrix} a\cos(\nu\theta) & 0 \\ 0 & a\sin(\nu\theta) \end{Bmatrix} \cdot \vec{H}^{(1)}(r,\nu,\hat{\omega}_{m\nu}) \cdot \vec{A}_m , \\ \vec{U}^{(2)}_{m\nu}(r,\theta) \\ \quad = \begin{Bmatrix} a\sin(\nu\theta) & 0 \\ 0 & a\cos(\nu\theta) \end{Bmatrix} \cdot \vec{H}^{(2)}(r,\nu,\hat{\omega}_{m\nu}) \cdot \vec{A}_m , \end{cases} \tag{10}$$

where

$$\vec{H}^{(i)}(r,\nu,\hat{\omega}_{m\nu}) = \begin{Bmatrix} e\hat{\omega}_{m\nu}J'_\nu(e\hat{\omega}_{m\nu}r) & (-1)^{i+1}\nu J_\nu(\hat{\omega}_{m\nu}r)/r \\ (-1)^i \nu J_\nu(e\hat{\omega}_{m\nu}r)/r & -\hat{\omega}_{m\nu}J'_\nu(\hat{\omega}_{m\nu}r) \end{Bmatrix}. \tag{11}$$

and $\hat{\omega}_{m\nu}=\omega_{m\nu}\sqrt{\rho/\mu}$, A_{m} satisfying an eigenfunction problem for each given ν with the same dispersion equation as Eq. (9). Here m represents the different branch of the dispersion curve. In addition, e is a coefficient related to the elastic constants by $e=\sqrt{\mu/(\lambda+2\mu)}$. The dispersion curves for the Rayleigh and first to tenth Whispering Gallery modes were obtained with the frequency value corresponding to each integer ν from 1 up to 400. Then, the summation over ν was carried out for each individual mode to get the time-dependent response of each mode. The Rayleigh, first and tenth Whispering Gallery modes (WG1,WG10) with corresponding m values of 1, 2, and 11 are shown in Fig. 6. Finally, the numerical wave form of each mode was superposed to obtain the total response. As shown in the top of Fig. 6, only the circumferentially reflected transverse waves ($2S$,$3S$) and their one-roundtrip sequences ($5S_1$,...,$11S_1$) are visible besides the Rayleigh mode. Here, the summations over ν and m are truncated at 400 and 11, respectively. Since the contribution of each mode is decreasing rapidly with the increase of ν and m, further increase of these two limits will not change the final wave form by more than 10%. Therefore, we conclude that the normal mode expansion method cannot reveal the same acoustic wave information as presented in Sec. IV. This method fails to solve the transient response of the cylindrical rod, even though it could predict the wave form of each mode.

VI. EXPERIMENTAL VALIDATION

In the previous sections, waves are identified on numerical results obtained by several methods. Although the numerical results obtained by the improved method are good evidence to support the wave identification, it is still necessary to confirm whether these waves exist or not in experimentally observed wave forms.

A Nd:YAG laser is used for ultrasonic wave generation in either the ablation or the thermoelastic regime. The pulse duration is 20 ns and the infrared light emission is obtained at 1064 nm with a maximum burst energy output of 340 mJ. The collimated optical beam is focused by means of a cylindrical lens (focus length is 150 mm). The line length and

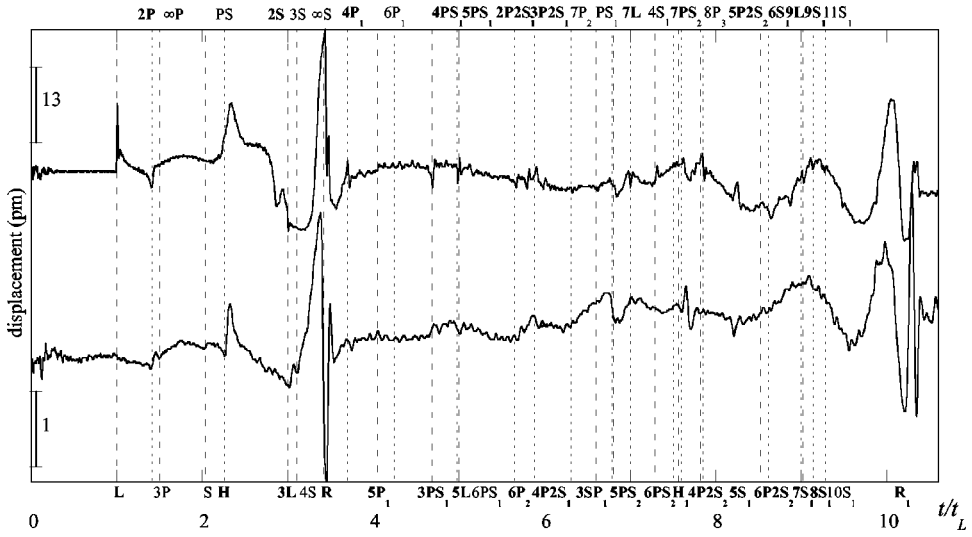


FIG. 7. Acoustic waves observed experimentally by the laser ultrasonic technique for aluminum cylinders at the epicenter of the laser line source generation under either the ablation (top) or the thermelastic (below) regime. Symbols at the top or bottom of each dashed line, respectively, denote the wave marked by the dense or sparse dashed line.

width are about 4 cm and 0.1 mm, respectively. Using an optical heterodyne probe with a power output of 100 mW and with a sensitivity[5] of $10^{-14}\,\mathrm{m}/\sqrt{\mathrm{Hz}}$, the radial displacement on the surface is detected at the epicenter direction $\theta=180°$. For the ablation generation, the signal was averaged by 15 shots; and for the thermoelastic generation, the signal was averaged by 110 shots. See Ref. 23 for the detailed description of the experimental setup.

Two aluminum cylinders were chosen for the experiments. Their diameters are $2a=9.97$ mm for the ablation regime and $2a=4.12$ mm for the thermoelastic regime. The longitudinal velocity was measured by the time arrivals of L, $3L$, $5L$, $7L$ as 6.36 ± 0.01 km/s, and the transverse wave velocity was measured by the average time arrivals of $2S$ and $5S_1$ as 3.13 ± 0.01 km/s.

First let us look at the experimental waves under the ablation regime. As shown in the top wave form of Fig. 7, the L, $3L$, $5L$, $7L$ are clearly visible. Wave L is positively polarized, $3L$ and $7L$ are negatively polarized. Wave $5L$ is positively polarized as it is separated with $4PS_1$. They are in accordance with that of the theoretical result in Fig. 1. Wave $9L$ is overlapped by another wave $8S_1$. Waves $2P$, $4P_1$, $6P_2$ are clearly observable. $8P_3$ is overlapped by another wave $4P2S_2$. Waves $2S$, $5S_1$,...,$9S_1$ are clearly visible, while $4S_1$ is hardly visible as predicted by theory. H is visible while PS is invisible, this reveals the effect of the beamwidth and time duration of the laser line source.[15] Wave PS_1 is naturally invisible, since it is the sequence of PS. Waves $3PS_1$, $4PS_1$, $5PS_1$ are observable. Further, $6PS_2$ exists, and $7PS_2$ coexists with H_1 because of the overlap. Waves $2P2S_1$, $3P2S_1$, $4P2S_1$ are also observable. Other waves appearing in the numerical results are hardly visible due to the effect of the physical dimension of the laser source.[15]

Now let us focus on the experimental waves generated under the thermoelastic regime. As shown in the low part of Fig. 7, this wave form is quite different from that for the ablation regime due to the differences in directivities. Here L, $3L$, $5L$, $7L$, $9L$ are no more observable. Waves $2P$, H, H_1, and $5S_1$ are quite clear, but other waves appearing in the theory are also hardly visible for a similar reason.

Although the shapes of Rayleigh waves R and R_1 in both regimes are not precisely coincident with that of the numerical results, the evolution of the shapes due to dispersion, the relative amplitudes, arrival times agree with the theoretical predictions. Specifically, the Rayleigh wave R is a monopolar in the experiment under the ablation regime, while it is a bipolar in the corresponding calculation in Fig. 1. For this experiment, the ratio of line source length to source–receiver distance is about 4, which may not be adequate to regard the laser line source as infinite.[24]

Finally let us underline that the theory has allowed us to fully identify the waves in the experimental displacement. The presented model along with the numerical inverse scheme and the ray trajectory analysis could also provide wave forms for any observation angle other than 180°, and the generated acoustic waves could also be identified.

VII. CONCLUSION

The boundary condition of our previous model on the 2D transient wave propagation in a homogeneous and isotropic cylinder was modified to avoid the numerical integration necessitated by the numerical inverse scheme. The Fourier series expansion was successfully applied to resolve this transient response problem, and only a discrete summation along wave number was needed for the inverse scheme. The numerical displacements in either the ablation or the thermoelastic regime are obtained with little numerical noise and short computation time. This improved method reveals more acoustic waves than the normal mode expansion method, which has been successfully applied to the 2D and 3D transient wave propagations in a hollow cylinder. Acoustic waves are fully identified by the ray trajectory analysis. These identified waves are further verified on the experimental results observed by the laser ultrasonic technique.

The identified rich responses are useful for the inverse problem of the nondestructive evaluation of cylindrical parts. For instance, waves such as the circumferentially reflected transverse waves ($2S$,$5S_1$) and the converted waves ($4PS_1$,$2P2S_1$,$3P2S_1$,$4P2S_1$) can be adopted for the accurate measurement of the longitudinal and transverse wave velocities for the elastic constant recovery[15] under ablation

generation. The modified physical model and the novel numerical inverse scheme will pave the way for the 3D transient wave propagation in a transversely isotropic cylinder generated by a laser point source.

ACKNOWLEDGMENTS

This work was done at Laboratoire de Mécanique Physique, Université Bordeaux 1. Y.P. was supported by KC-Wong and CNRS fellowship and by Natural Science Foundation of China under Grant No. 10234060.

[1] K. Sezawa, ''Dispersion of elastic waves propagaed on the surface of stratified bodies and on curved surfaces'' Bull. Earthquake Res. Inst., Univ. Tokyo **3**, 1–18 (1927).
[2] A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[3] H. Uberall, *Physical Acoustics* (Academic, New York, 1973), Vol. 10, pp. 1–60.
[4] C. B. Scruby and L. E. Drain, *Laser Ultrasonics: Techniques and Applications* (Hilger, New York, 1990).
[5] D. Royer, E. Dieulesaint, X. Jia, and Y. Shui, ''Optical generation and detection of surface acoustic waves on a sphere,'' Appl. Phys. Lett. **52**, 706–708 (1988).
[6] S. Ishikawa, H. Cho, Y. Tsukahara, N. Nakaso, and K. Yamanaka, ''Analysis of spurious bulk waves in ball surface wave device,'' Ultrasonics **41**, 1–8 (2003).
[7] M. Qian and X. Wu, ''Studies of cylindrical Rayleigh waves with laser ultrasonics,'' Prog. Nat. Sci. **11**, s258–s264 (2001).
[8] D. Clorennec and D. Royer, ''Analysis of surface acoustic wave propagation on a cylinder using laser ultrasonics,'' Appl. Phys. Lett. **82**, 4608–4610 (2003).
[9] X. Wu and M. Qian, ''Simulation of the finite element method on wave propagation in cylinders,'' Prog. Nat. Sci. **11**, s265–s268 (2001).
[10] G. Liu and J. Qu, ''Transient wave propagation in a circular annulus subjected to transient excitation on its outer surface,'' J. Acoust. Soc. Am. **104**, 1210–1220 (1998).
[11] L. Tang and J. Cheng, ''Numerical analysis on laser-generated guided elastic waves in a hollow cylinder,'' J. Nondestruct. Eval. **21**, 45–53 (2002).
[12] M. Qian, M. Duquennoy, M. Ouatouh, F. Jenot, and M. Ourak, ''Laser ultrasonic characterization of surface residual stresses in steel rods,'' Rev. Prog. Quant. Nondestr. Eval. **20**, 1487–1493 (2001).
[13] D. Clorennec, D. Royer, and H. Walaszek, ''Nondestructive evaluation of cylindrical parts using laser ultrasonics,'' Ultrasonics **40**, 783–789 (2002).
[14] Y. Pan, C. Rossignol, and B. Audoin, ''Acoustic waves generated by a laser line pulse in a transversely isotropic cylinder,'' Appl. Phys. Lett. **82**, 4379–4381 (2003).
[15] Y. Pan, C. Rossignol, and B. Audoin, ''Acoustic waves generated by a laser line pulse in cylinders: Application to the elastic constants measurement,'' J. Acoust. Soc. Am. **115**, 1537–1545 (2004).
[16] Y. Pan, C. Rossignol, and B. Audoin, ''Acoustic waves generated by a laser point source in an isotropic cylinder,'' J. Acoust. Soc. Am. **116**, 814–820 (2004).
[17] K. Aki and P. G. Richards, *Quantitative Seismology Theory and Methods* (Freeman, New York, 1943).
[18] D. A. Hutchins, *Physical Acoustics* (Academic, New York, 1988), Vol. 18, pp. 21–123.
[19] L. R. F. Rose, ''Point-source representation for laser-generated ultrasound,'' J. Acoust. Soc. Am. **75**, 723–732 (1984).
[20] C. B. Scruby, R. J. Dewhurst, D. A. Hutchins, and S. B. Palmer, ''Quantitative studies of thermally generated elastic waves in laser-irradiated metals,'' J. Appl. Phys. **51**, 6210–6216 (1980).
[21] X. Wu, M. Qian, and J. H. Cantrell, ''Dispersive properties of cylindrical waves,'' Appl. Phys. Lett. **83**, 4053–4055 (2003).
[22] J. D. Aussel, A. Le Brun, and J. C. Baboux, ''Generating acoustic waves by laser: Theoretical and experimental study of the emission source,'' Ultrasonics **26**, 245–255 (1988).
[23] F. Reverdy and B. Audoin, ''Elastic constants determination of anisotropic materials from phase velocities of acoustic waves generated and detected by lasers,'' J. Acoust. Soc. Am. **109**, 1965–1972 (2001).
[24] D. Royer and C. Chenu, ''Experimental and theoretical waveforms of Rayleigh waves generated by a thermoelastic laser line source,'' Ultrasonics **38**, 891–895 (2000).

Shear-horizontal acoustic wave propagation in piezoelectric bounded plates with metal gratings

Shi Chen, Tiantong Tang, and Zhaohong Wang
Telecom College, Xi'an Jiaotong University, Xi'an 710049, China

(Received 1 October 2004; revised 3 March 2005; accepted 4 March 2005)

In this paper, shear-horizontal (SH) acoustic wave propagation in metal gratings deposited on piezoelectric bounded plates is investigated. The spectral characteristics of the electromechanical coupling coefficient are studied first, which are very important for acoustic wave device designs. And, an effective mathematic method based on even- and odd base functions is also presented for overcoming the large frequency thickness product problem. Then, the characteristics of the grating modes are studied, and the nature and characteristics of the stop bands are investigated fully. The results show that the width and attenuation of the stop bands are dominated by the electromechanical coupling coefficient at the frequency centers of the stop bands. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1898083]

PACS numbers: 43.35.Ns, 43.35.Pt, 43.35.Cg [TDM] Pages: 3609–3615

I. INTRODUCTION

Metal grating structure components are widely used in modern electronic devices and instruments, such as filters, delay lines, interdigital transducers, sensors, and mobile phone systems. So, it is necessary to understand fully their characteristics of wave propagation to improve the performance of the devices.

The characteristics of the acoustic wave propagation in metal grating structures have been studied extensively by many authors. Surface acoustic waves, leaky surface acoustic waves, and longitudinal leaky surface acoustic waves have been studied using boundary integral method and Floquet theorem.[1,2] Other methods, including finite Element method with special periodic boundary conditions,[3] methods based on periodic Green's function defined on the interface between piezoelectrics and free space,[4] and that based on Legendre polynomial expansion of electric field[5] have been used as well for these problems.

Multilayer structures have given rise to considerable attention to designs of many modern electronic devices and instruments. Acoustic wave propagation characteristics in an infinite piezoelectric plate were studied by Tiersten[6] and Bleustein[7] in 1963 and 1969, respectively, and it was discovered that shear-horizontal (SH) acoustic waves can exist in the structures. Because of the increasing need for new communication device designs, many kinds of multilayer structures have been extensively investigated recently by many authors,[8–12] and new devices have been developed based on them.

A theoretical formulation for the SH wave excitation in piezoelectric bounded plates by use of interdigital transducers(IDTs) was presented by Wang and Varadan in 2002,[13] and the SH wave propagation at the electric short-circuit condition in the same structures was also studied by the same authors.[14] However, the analytical method in the literature[14] is subject to serious drawbacks (which are discussed in detail in Sec. II) when the large frequency thickness product ωh (Refs. 15, 16) problem is encountered, and it becomes very complicated and even invalid when the number of layers is large (such as if the number of layers is larger than 4). The electromechanical coupling coefficient (ECC) is not investigated, which is a crucial parameter of the IDTs and controls the excitation efficiency of acoustic waves. Moreover, the characteristics of SH wave propagation on metal gratings deposited on piezoelectric bounded plates, which are important in the designs of many acoustic wave devices, are not studied.

In Sec. II of this paper, an effective analytical method based on even- and odd base functions is presented to overcome the drawbacks mentioned previously, which can be easily extended to many multilayer structure problems by using the rigorous symbol calculation functions(for linear matrix equation) in MATLAB software. And, the ECC of the plate modes (the wave modes in piezoelectric bounded plates without gratings) is investigated. By using this method the performances of the devices based on the structures are analyzed. In Sec. III, the characteristics of the grating modes (the wave modes in the gratings) are studied based on a boundary integral equation and Floquet theorem. It is found that the width and attenuation of the stop bands are dominated by the ECC at the frequency centers of the stop bands. Compared with the conventional study methods[6,7,14] focused on the field distribution of modes, the method in this paper regarding the ECC reveals more information on the performance of the devices.

In numerical calculation, the anisotropic piezoelectric materials with point-group symmetry $6mm$ class are considered.

II. SH ACOUSTIC WAVES IN PIEZOELECTRIC BOUNDED PLATES

In Fig. 1 the structure of a piezoelectric bounded plate for theoretical analysis is shown. It is infinite in $x-z$ plane; thicknesses of the piezoelectric and metal layers are denoted by $2h_p$ and $2h_m$, respectively.

Because the structure considered here is infinite in the x- and z directions, choosing x as the wave propagating direction, all the field variables are independent on the z coordi-

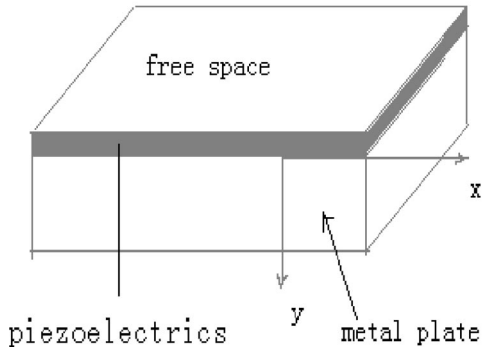


FIG. 1. An infinite metal plate bounded by a layer of piezoelectric material.

nate, i.e., $\partial/\partial z=0$. For the SH acoustic waves, only the z direction particle displacement u_3 is nonzero and u_3 and the electric potential ϕ are chosen as independent variables in the acoustic field analysis hereafter.

Using the quasistatic approximation, the electric field equation in the free space is

$$\nabla_f^2\phi^f=0. \tag{1}$$

The field equations for the particle displacement u_3 and electric potential in the piezoelectrics are[14]

$$c_{44}\nabla_p^2u_3^p+e_{15}\nabla_p^2\phi^p=\rho\frac{\partial^2}{\partial t^2}u_3^p, \tag{2}$$

$$e_{15}\nabla_p^2u_3^p-\varepsilon_{11}\nabla_p^2\phi^p=0. \tag{3}$$

In the metal layer, there is no electric field and the field equation is only for the particle displacement u_3

$$c_{44}'\nabla_m^2u_3^m=\rho'\frac{\partial^2}{\partial t^2}u_3^m. \tag{4}$$

In these equations, superscripts f, p, and m denote fields in the free space, the piezoelectrics and the metal material, respectively, c_{44} and c_{44}' are the components of the elastic tensor in the piezoelectrics and the metal material, respectively, e_{15} is the component of the piezoelectric tensor in the piezoelectrics, ρ and ρ' are the mass density of the piezoelectrics and the metal material, respectively. There local coordinate systems are adopted for any layer due to the cause to be discussed afterward, i.e. (x,y_ζ,z) for the ζ layer, where $\zeta=f,p,m$, which denote the free space, the piezoelectric layer, and the metal layer, respectively. The origins of the axes y_p and y_m are on the thickness centers of the piezoelectrics and metal layers, respectively. The origin of the axis y_f is on the interface between the free space and the piezoelectrics layer. And, $\nabla_\zeta^2=(\partial^2/\partial x^2)+(\partial^2/\partial y_\zeta^2)$.

Assuming that the waves are monochromatic and propagate in the x direction, then all field variables have a common factor $\exp(-j\omega t)\exp(jk_xx)$, where ω denotes the angular frequency and k_x denotes the wave number. General solutions for Eqs. (1)–(4) can be found in the literature[14] and are listed as follows for discussion:

$$\phi^f=A'\exp(|k_x|y_f)\exp(jk_xx), \tag{5}$$

$$u_3^p=[B'\exp(k_py_p)+C'\exp(-k_py_p)]\exp(jk_xx), \tag{6}$$

$$\phi^p=\Big\{D'\exp(k_xy_p)+E'\exp(-k_xy_p)+\frac{e_{15}}{\varepsilon_{11}}[B'\exp(k_py_p)+C'\times\exp(-k_py_p)]\Big\}\exp(jk_xx), \tag{7}$$

$$u_3^m=[F'\exp(k_my_m)+G'\exp(-k_my_m)]\exp(jk_xx), \tag{8}$$

where $k_p=\sqrt{k_x^2-[(\rho\omega^2\varepsilon_{11})/(c_{44}\varepsilon_{11}+e_{15}^2)]}$, $k_m=\sqrt{k_x^2-\rho'\omega^2/c_{44}'}$.

Because the field solutions [in Eqs. (5)–(8)] depend on the y_ζ coordinate in exponent form, when the partial waves attenuate or increase in the y_ζ direction [implying that $real(k_\zeta)\neq0$], and the frequency thickness product ωh (which is proportional to the $k_\zeta h_\zeta$) becomes extremely large, in numerical calculation the overflow and low precision may be caused by the extremely large ωh.[15,16] This conclusion can also be testified to by observing Eqs. (37) and (48) in Ref. 14. The overflow may be overcome simply by dividing the two sides of the equation (where the overflow emerges) by a large exponent $\exp(k_\zeta h_\zeta)$ [where $real(k_\zeta h_\zeta)>0$]. However, it is not wise to do it this way when many layers are considered, because it will lead to the loss of the precision or new difficulties regarding order matching. In order to avoid these problems, new even- and odd base functions are adopted in this paper. Then, the general solutions can be expressed as follows:

$$\phi^f=[A\varphi_f^S(k_x,y_f)+A\varphi_f^A(k_x,y_f)]\exp(jk_xx), \tag{9}$$

$$u_3^p=[B\psi_p^S(k_p,y_p)+C\psi_p^A(k_p,y_p)]\exp(jk_xx), \tag{10}$$

$$\phi^p=\Big\{D\varphi_p^S(k_x,y_p)+E\varphi_p^A(k_x,y_p)+\frac{e_{15}}{\varepsilon_{11}}[B\psi_p^S(k_p,y_p)+C\psi_p^A(k_p,y_p)]\Big\}\exp(jk_xx), \tag{11}$$

$$u_3^m=[F\psi_m^S(k_m,y_m)+G\psi_m^A(k_m,y_m)]\exp(jk_xx), \tag{12}$$

where

$$\psi_\zeta^S(k_\zeta,y_\zeta)=\frac{\exp(k_\zeta y_\zeta)+\exp(-k_\zeta y_\zeta)}{\exp(k_\zeta h_\zeta)+\exp(-k_\zeta h_\zeta)},$$

$$\psi_\zeta^A(k_\zeta,y_\zeta)=\frac{\exp(k_\zeta y_\zeta)-\exp(-k_\zeta y_\zeta)}{\exp(k_\zeta h_\zeta)+\exp(-k_\zeta h_\zeta)}, \tag{13}$$

$$\varphi_\zeta^S(k_x,y_\zeta)=\frac{\exp(k_xy_\zeta)+\exp(-k_xy_\zeta)}{\exp(k_xh_\zeta)+\exp(-k_xh_\zeta)},$$

$$\varphi_\zeta^A(k_x,y_\zeta)=\frac{\exp(k_xy_\zeta)-\exp(-k_xy_\zeta)}{\exp(k_xh_\zeta)+\exp(-k_xh_\zeta)}, \tag{14}$$

where $h_f=0$, $\psi_\zeta^S(k_\zeta,y_\zeta)$, $\varphi_\zeta^S(k_x,y_\zeta)$, $\psi_\zeta^A(k_\zeta,y_\zeta)$, and $\varphi_\zeta^A(k_x,y_\zeta)$ are even- and odd base functions. It is easily checked that Eqs. (9)–(12) are general solutions of Eqs. (1)–(4). If the k_ζ and k_x are chosen according to the following principle, i.e, $real(k_\zeta)\geqslant0$ and $real(k_x)\geqslant0$, then expressions (13)–(14) can be rewritten as follows:

$$\psi_\zeta^S(k_\zeta,y_\zeta)=\frac{\exp[-k_\zeta(h_\zeta-y_\zeta)]+\exp[-k_\zeta(h_\zeta+y_\zeta)]}{1+\exp(-2k_\zeta h_\zeta)},$$

$$\psi_\zeta^A(k_\zeta,y_\zeta)=\frac{\exp[-k_\zeta(h_\zeta-y_\zeta)]-\exp[-k_\zeta(h_\zeta+y_\zeta)]}{1+\exp(-2k_\zeta h_\zeta)},$$

$$\varphi_\zeta^S(k_x,y_\zeta)=\frac{\exp[-k_x(h_\zeta-y_\zeta)]+\exp[-k_x(h_\zeta+y_\zeta)]}{1+\exp(-2k_x h_\zeta)},$$

$$\varphi_\zeta^A(k_x,y_\zeta)=\frac{\exp[-k_x(h_\zeta-y_\zeta)]-\exp[-k_x(h_\zeta+y_\zeta)]}{1+\exp(-2k_x h_\zeta)}, \tag{15}$$

because $h_\zeta-y_\zeta\geqslant 0$ and $h_\zeta+y_\zeta\geqslant 0$ in the $\zeta(\zeta=p,m)$ layer, so the modules of the exponential items appearing in the expressions (15) are all less than 1, and the orders of the modules of $\psi_\zeta^S(k_\zeta,y_\zeta)$, $\varphi_\zeta^S(k_x,y_\zeta)$, $\psi_\zeta^A(k_\zeta,y_\zeta)$, and $\varphi_\zeta^A(k_x,y_\zeta)$ are the same and all small. These imply that all the coefficients of boundary condition equations are balanced automatically, so the high precision and the numerical stability are guaranteed and the overflow is overcome completely in the numerical calculation.

Equation (9) implies that the electric potential should be attenuated to zero when $y_f\to-\infty$. In order to obtain the characteristics of the SH wave propagation, the proper boundary conditions must be considered; this knowledge can be found in the literature.[14] So, only main steps and results are presented.

As a sample, the case of the electric open-circuit condition is considered here. In the following derivation B is chosen as the independent variable. From the boundary conditions except for the electric displacement continuous boundary condition on the interface between the free space and the piezoelectrics, the variables A, C, D, E, F, G can be expressed as functions of B, k_x, and ω. The derivation is direct and simple, and for multilayer systems (the number of layers is larger than 4) the rigorous symbol calculation functions (for the matrix equation) of the MATLAB software can be utilized. This will simplify the tedious derivation drastically and will make it possible to extend the method to many relative fields. The results are listed directly as follows:

$$C=C(k_x,\omega)B$$
$$=\frac{-b\varphi_p^A+d\psi_p^A-(e/a)(1+\varphi_p^A\varphi_p^A)}{d+b\psi_p^A\varphi_p^A-[d-(e/a)\psi_p^A](1+\varphi_p^A\varphi_p^A)}B, \tag{16}$$

$$G=G(k_x,\omega)B=[1/a+(\psi_p^A/a)C(k_x,\omega)]B, \tag{17}$$

$$F=F(k_x,\omega)B=-(1/\psi_m^A)G(k_x,\omega)B, \tag{18}$$

$$E=E(k_x,\omega)B=[eG(k_x,\omega)-dC(k_x,\omega)]B, \tag{19}$$

$$D=D(k_x,\omega)B=[-d\psi_p^A/\varphi_p^A+(e/\varphi_p^A)G(k_x,\omega)]B, \tag{20}$$

$$A=A(k_x,\omega)B=[2b+2D(k_x,\omega)]B, \tag{21}$$

where $\psi_\zeta^A\equiv\psi_\zeta^A(k_\zeta,h_\zeta)$, $\varphi_\zeta^A\equiv\varphi_\zeta^A(k_x,h_\zeta)$, $a=-\psi_m^A-1/\psi_m^A$, $b=e_{15}/\varepsilon_{11}$, $d=(c_{44}/e_{15}+e_{15}/\varepsilon_{11})(k_p/k_x)$, $e=c'_{44}k_m/(e_{15}k_x)$. In order to obtain the dispersion equation, the electric displacement continuous boundary condition on the interface between the free space and the piezoelectrics must be satisfied

$$[\varepsilon_0 A(k_x,\omega)-\varepsilon_{11}E(k_x,\omega)+\varepsilon_{11}\varphi_p^A D(k_x,\omega)]B=0. \tag{22}$$

In order to obtain a nontrivial solution, the following equation must be satisfied:

$$\varepsilon_0 A(k_x,\omega)-\varepsilon_{11}E(k_x,\omega)+\varepsilon_{11}\varphi_p^A D(k_x,\omega)=0. \tag{23}$$

Equation (23) is the dispersion equation for the case of the electric open-circuit condition. The dispersion equation for the case of the electric short-circuit condition can be obtained as well by using a similar procedure.

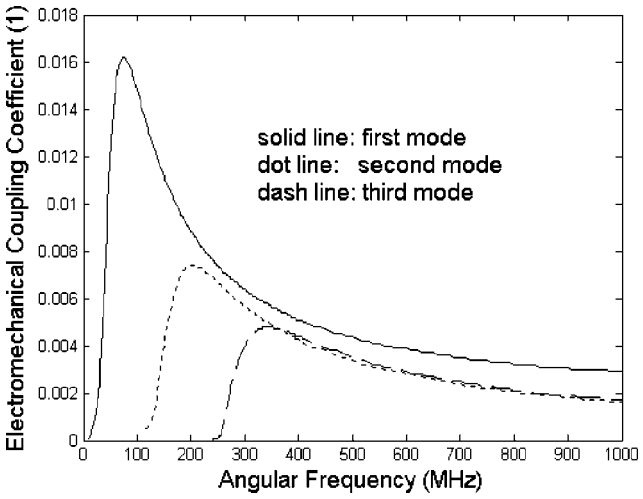


FIG. 2. The κ of the plate modes in the ZnO-steel plate with $h_p=0.05$(mm) and $h_m=0.25$(mm).

In Fig. 2 the computed electromechanical coupling coefficient κ for different modes as functions of the angular frequency is presented. Here, κ is defined as $\kappa=(v_f-v_m)/v_f$, where v_f and v_m are the phase velocities at the electric open-circuit condition and the electric short-circuit condition, respectively. It is found that the κ of the first mode is much larger than that of higher order modes. The reason may be that the first mode is a surface acoustic wave[14] which is easily affected by the surface electric boundary conditions, and higher order modes are not confined to the surface[14] and are not easily affected by the surface electric boundary conditions. At the very high-frequency regime, the κ of all the modes becomes very small. These results imply that only surface waves (at the low-frequency regime) can be effectively excited and made use of in the structures.

In Fig. 3 and Fig. 4 the spectral characteristics of the κ of the first mode for different piezoelectric and metal layer thickness ratios (h_p/h_m) are shown. When the thickness of the piezoelectric layer is fixed (in Fig. 3, $2h_p=0.1$ mm), it is observed that the κ for different ratios tend to be identical at higher frequency regime and for smaller thickness ratio there is a higher peak. The curves of the κ for smaller ratios ($h_p/h_m<0.1$) are identical with the dotted line in Fig. 3; this is because the surface waves are not affected by the variation of the h_m when the h_m is very large. And, the curves of the κ for larger ratios ($h_p/h_m>10$) are identical to the dash-dot line in Fig. 3, because the mass loading effects of the metal

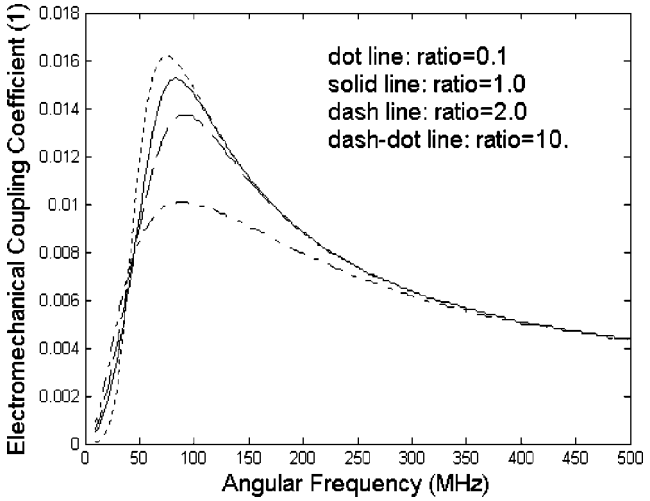


FIG. 3. The κ of the first plate modes in the ZnO-steel plate with different (h_p/h_m) ratios (where $h_p=0.05$ mm).

layer can be ignored when the h_m is very small (for the case the wave modes have been studied in the literature[6,7]). These results show that, compared with a sole piezoelectric plate,[6,7] the combination of piezoelectric layer and metal plate can improve the excitation efficiency of SH surface waves.

When the thickness of the metal layer is fixed (in Fig. 4, $2h_m=1$ mm), it is observed that the peak of the κ shifts to high frequency and the width of the peak becomes larger with the decreasing of the ratio (h_p/h_m). These bring great flexibility to the device designs.

In all the numerical calculation, the method in this paper is identical with the method in the literature[14] when the ωh is small. However, the method in this paper can be used to deal with the problems with large ωh, and it can be used also to calculate the SH modes of multilayer structures (where the number of layers is larger than 4).

III. SH ACOUSTIC WAVES IN PIEZOELECTRIC BOUNDED PLATES WITH METAL GRATINGS

An infinite piezoelectric bounded plate with a metal grating is shown in Fig. 5. In the figure $2h_p$ and $2h_m$ denote

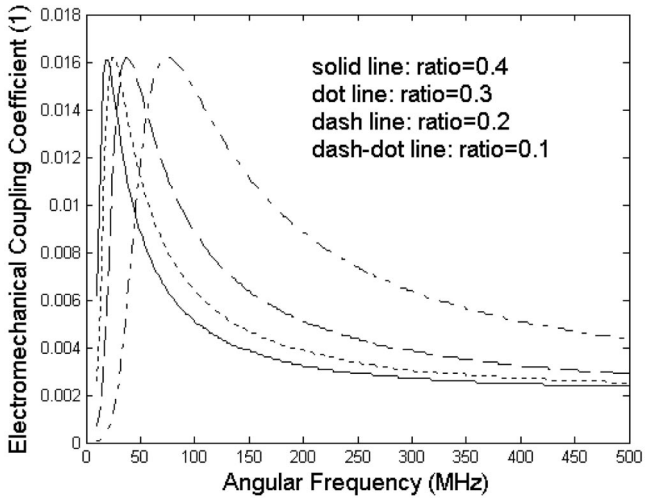


FIG. 4. The κ of the first plate modes in the ZnO-steel plate with different (h_p/h_m) ratios (where $h_m=0.5$ mm).

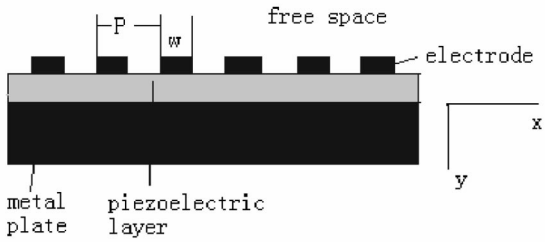


FIG. 5. An infinite metal grating deposited on piezoelectric bounded plates.

the thicknesses of the piezoelectric and metal layer, respectively, P denotes the period of the grating, and w denotes the width of a grating electrode.

The field components of the periodic system can be expressed using Floquet theorem as follows:

$$\phi^f=\sum_{n=-\infty}^{\infty} A_n[\varphi_f^S(\beta_n,y_f)+\varphi_f^A(\beta_n,y_f)]\exp(j\beta_n x), \tag{24}$$

$$u_3^p=\sum_{n=-\infty}^{\infty}[B_n\psi_p^S(k_{p,n},y_p) + C_n\psi_p^A(k_{p,n},y_p)]\exp(j\beta_n x), \tag{25}$$

$$\phi^p=\sum_{n=-\infty}^{\infty}\left\{D_n\varphi_p^S(\beta_n,y_p)+E_n\varphi_p^A(\beta_n,y_p) + \frac{e_{15}}{\varepsilon_{11}}[B_n\psi_p^S(k_{p,n},y_p) + C_n\psi_p^A(k_{p,n},y_p)]\right\}\exp(j\beta_n x), \tag{26}$$

$$u_3^m=\sum_{n=-\infty}^{\infty}[F_n\psi_m^S(k_{m,n},y_m) + G_n\psi_m^A(k_{m,n},y_m)]\exp(j\beta_n x), \tag{27}$$

where $k_{p,n}=\sqrt{\beta_n^2-\rho\omega^2\varepsilon_{11}/(c_{44}\varepsilon_{11}+e_{15}^2)}$, $k_{m,n}=\sqrt{\beta_n^2-(\rho'\omega^2/c_{44}')}$, $\beta_n=(2\pi/P)n+\gamma$, where n is a integer number, and γ is a complex wave number of the fundamental space harmonic whose real and imaginary parts are referred to as the grating velocity and the attenuation of the wave respectively. The grating velocity of the grating modes can be expressed as $v_{\text{grating}}=\omega/\text{real}(\gamma)$ and is confined to the first Brillouin zone.

Assuming that the thicknesses of the electrodes are infinitely thin, then the mass loading effects of electrodes can be ignored. It is obvious that Eqs. (24)–(27) are the general solutions of Eqs. (1)–(4). One needs only to find the amplitude coefficients of all harmonics by using the proper boundary conditions.[14]

For any harmonic, as mentioned previously, all the amplitude coefficients A_n, B_n, C_n, D_n, E_n, F_n, G_n can be expressed as functions of B_n, ω, and γ. The following expressions can be obtained according to expressions (16)–(21):

$$A_n = A(\beta_n, \omega)B_n, \quad C_n = C(\beta_n, \omega)B_n,$$

$$D_n = D(\beta_n, \omega)B_n,$$

$$E_n = E(\beta_n, \omega)B_n, \quad F_n = F(\beta_n, \omega)B_n,$$

$$G_n = G(\beta_n, \omega)B_n. \tag{28}$$

Expressions (28) imply that the boundary conditions except for the electric boundary conditions on the interface between the free space and the piezoelectrics are satisfied. In order to get B_n, the electric boundary conditions must be considered, i.e.,

$$\phi^f = 0 \quad (\text{for the electrode region } 0 \leqslant x \leqslant w), \tag{29}$$

$$D^f = D^p \quad (\text{for the outside-electrode region } w \leqslant x \leqslant P), \tag{30}$$

for the short-circuit gratings, and

$$\phi^f = V_0 \quad (\text{for the electrode region } 0 \leqslant x \leqslant w), \tag{31}$$

$$D^f = D^p \quad (\text{for the outside-electrode region } w \leqslant x \leqslant P), \tag{32}$$

$$\int_0^w (D^p - D^f)dx = 0 \quad (\text{for the electrode region } 0 \leqslant x \leqslant w) \tag{33}$$

for the open-circuit gratings. Here, D^f and D^p denote the normal electric displacements in the free space and piezoelectrics, respectively. It is implicit from Eq. (33) that the total electric charge on a separated electrode must be zero.

Using expressions (28), the expressions of the electric potential and the electric displacement on the interface between free space and piezoelectrics are derived as follows:

$$\phi^f = \sum_{n=-\infty}^{\infty} \phi_n^f B_n \exp(j\beta_n x),$$

$$D^p = \sum_{n=-\infty}^{\infty} D_n^p B_n \exp(j\beta_n x),$$

$$D^f = \sum_{n=-\infty}^{\infty} D_n^f B_n \exp(j\beta_n x), \tag{34}$$

where ϕ_n^f, D_n^p, D_n^f are

$$\varphi_n^f = A(\beta_n, \omega), \quad D_n^f = -\varepsilon_0 \beta_n A(\beta_n, \omega),$$

$$D_n^p = -\varepsilon_{11}\beta_n[-\varphi_p^A(\beta_n, h_p)D(\beta_n, \omega) + E(\beta_n, \omega)]. \tag{35}$$

The Galerkin method is adopted to solve the problem. For the short-circuit gratings we have following equations:

$$\int_0^w \phi^f \exp(-j\beta_m x)dx + q\int_w^p (D^p - D^f)\exp(-j\beta_m x)dx = 0 \quad (m = 0, \pm 1, \pm 2, \pm 3\cdots), \tag{36}$$

and for the open-circuit gratings, we have the following equations

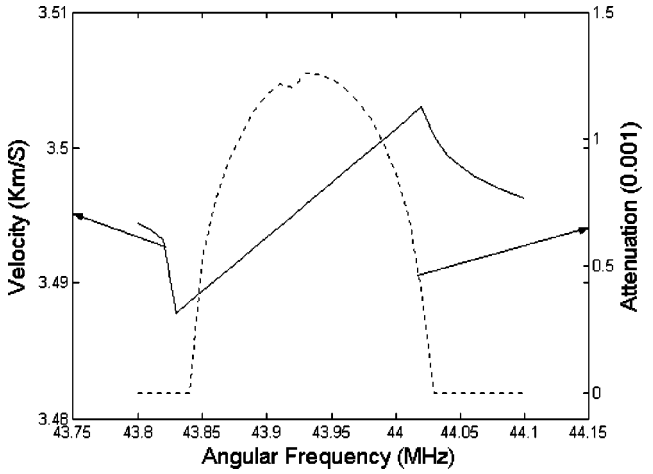


FIG. 6. The first stop band of the ZnO-steel plate grating with $h_p = 0.05$(mm), $h_m = 0.25$(mm), $p = 0.25$(mm), and $w = 0.125$(mm).

$$\int_0^w (\phi^f - V_0)\exp(-j\beta_m x)dx + q\int_w^p (D^p - D^f) \times \exp(-j\beta_m x)dx = 0 \quad (m = 0, \pm 1, \pm 2, \pm 3\cdots), \tag{37}$$

$$\int_0^w (D^p - D^f)dx = 0, \tag{38}$$

where q is a constant which makes the orders of ϕ^p, D^p, and D^f identical.

To carry out numerical calculation, harmonics with very high order have to be neglected. For the short-circuit gratings, considering only $-M, \ldots, -1, 0, 1, \ldots, M$ order harmonics, the final equation is

$$RT = 0, \tag{39}$$

where R is a $(2M+1)\times(2M+1)$ matrix, $T \equiv [T_1, T_2, \ldots, T_{2M+1}]^T = [B_{-M}, B_{-M+1}, \ldots, B_M]^T$, where the superscript T denotes the transpose operator. The components of the matrix R are

$$R_{m,n} = \int_0^w \phi_{n-1-M}^f \exp[j(\beta_{n-1-M} - \beta_{m-1-M})x]dx + q\int_w^p (D_{n-1-M}^p - D_{n-1-M}^f) \times \exp[j(\beta_{n-1-M} - \beta_{m-1-M})x]dx, \tag{40}$$

where $m, n = 1, 2, \ldots, 2M+1$.

For the open-circuit gratings, considering only $-M, \ldots, -1, 0, 1, \ldots, M-1$ order harmonics, the final equation is

$$R'T' = 0, \tag{41}$$

where R' is a $(2M+1)\times(2M+1)$ matrix, $T' \equiv [T_1', T_2', \ldots, T_{2M+1}']^T = [B_{-M}, B_{-M+1}, \ldots, B_{M-1}, V_0]^T$. The components of the matrix R' are

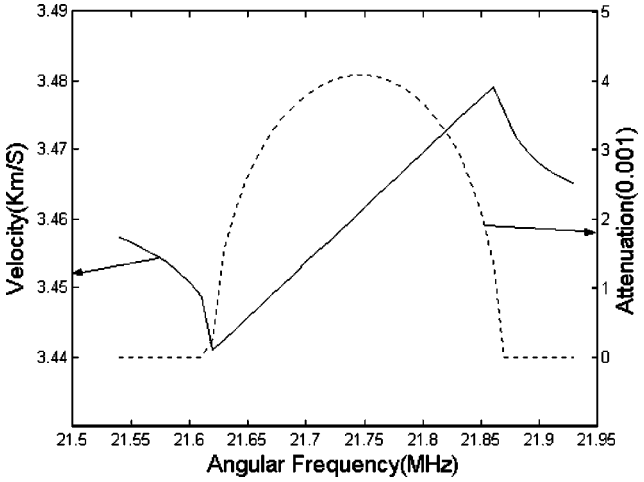


FIG. 7. The first stop band of the PZT4-steel plate grating with $h_p = 0.05$(mm), $h_m = 0.25$(mm), $p = 0.25$(mm), and $w = 0.125$(mm).

$$R'_{m,n} = R_{m,n},$$

$$R'_{m,2M+1} = \int_0^w -V_0 \exp(-j\beta_m x)\,dx,$$

$$R'_{2M+1,n} = \int_0^w (D_n^p - D_n^f)\exp(j\beta_{n-1-M}x)\,dx,$$

$$R'_{2M+1,2M+1} = 0, \tag{42}$$

where $m,n = 1,2,...,2M$.

In order to obtain nontrivial solutions, the determinants of R (for the short-circuit grating) and R' (for the open-circuit grating) must be zero.

The characteristics of the first stop band for the first mode in the gratings are investigated numerically by the previous method. In the numerical calculation, a good precision is obtained when $M = 5$. Compared with the method in the literature,[5] the computation time is much shorter because only a few harmonics are involved, and more complex grating structures can be dealt with by the method in this paper. But, the drawback of the method is that it cannot be used to deal with the excitation problems of the IDTs due to the singularity of the electric charge at the edge of electrodes.

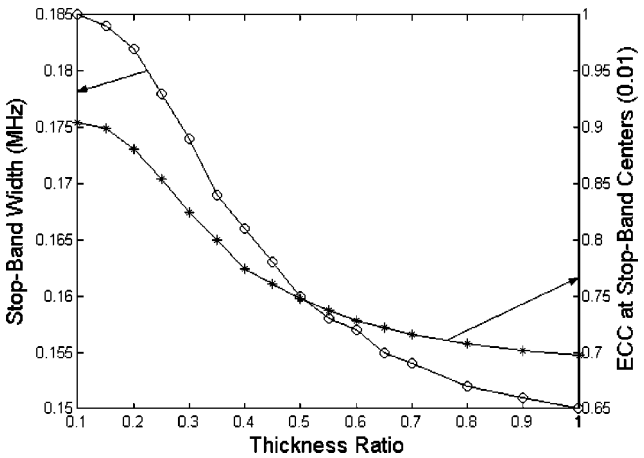


FIG. 8. The width of the stop bands and the electromechanical coupling coefficient (ECC) at the frequency centers of the stop bands with different thickness ratios (h_p/h_m), where $h_p = 0.05$(mm), $p = 0.25$(mm), and $w = 0.125$(mm).

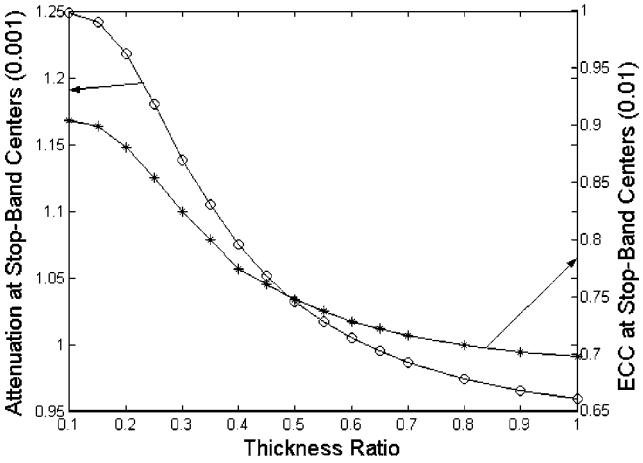


FIG. 9. The attenuation and the ECC at the frequency centers of the stop bands with different thickness ratios (h_p/h_m), where $h_p = 0.05$(mm), $p = 0.25$(mm), and $w = 0.125$(mm).

In Fig. 6 and Fig. 7 the dispersion characteristics are shown for the first stop band of the shorted circuit gratings, where the solid lines denote the grating velocity of the grating modes; the dotted lines denote the imaginary part of γ. Larger attenuation is found in the PZT4-steel Palte structure, because there is a larger ECC in the structure and the surface waves are affected more easily by the surface electric boundary conditions. For the frequency centers of the stop bands, the known Bragg reflection condition[17] is satisfied: $2p = n\lambda_s$, where $\lambda_s = 2\pi/real(\gamma)$ and $n = \pm 1, \pm 2, \pm 3...$, which reveals that the stop bands are caused by the surface waves reflection.

The characteristics of the stop bands are mainly described by the two parameters, i.e., the bandwidth and the largest attenuation of the stop bands. In order to reveal the nature and characteristics of the stop bands, the relations between the two parameters and the κ at the frequency centers of the stop bands are investigated and shown in Figs. 8, 9, 10, and 11, where only the first stop bands for the short-circuit gratings (on ZnO-steel plate) are studied (because

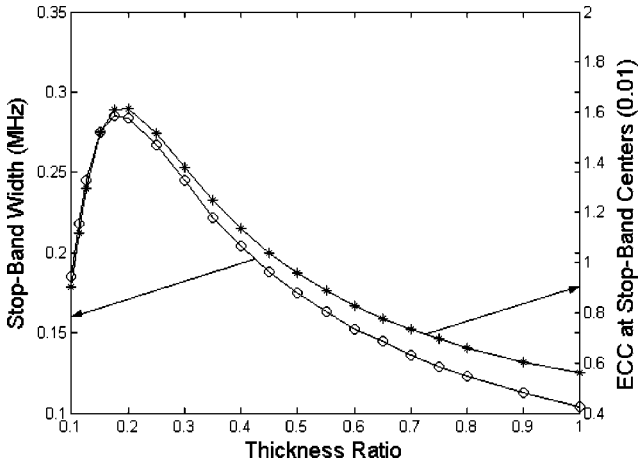


FIG. 10. The width of the stop bands and the ECC at the frequency centers of the stop bands with different thickness ratios (h_p/h_m), where $h_p = 0.5$(mm), $p = 0.25$(mm), and $w = 0.125$(mm).

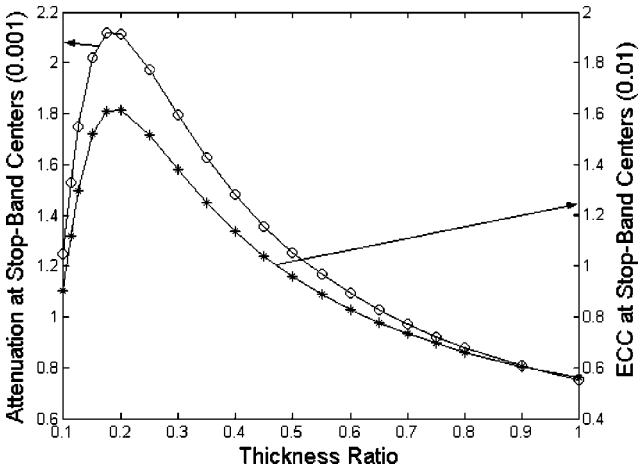


FIG. 11. The attenuation and the ECC at the frequency centers of the stop bands with different thickness ratios (h_p/h_m), where $h_m=0.5$(mm), $p=0.25$(mm), and $w=0.125$(mm).

there are the same characteristics for the open-circuit gratings). In Figs. 8 and 9, the gratings with the fixed h_p and different h_p/h_m ratios are considered, and in Figs. 10 and 11, the gratings with the fixed h_m and different h_p/h_m ratios are considered.

It is observed that the curves of the bandwidth and the largest attenuation (for the studied case, the attenuation at the frequency centers of the stop bands) of the stop bands has the same characteristics, with the curves of the κ at the frequency centers of the stop bands. In other words, the largest attenuation and the width of the stop bands are dominated by the κ, and there are larger attenuation and bandwidth of the stop bands for larger κ. It can be understood easily by considering the following extreme case, if $\kappa \to 0$, which implies that the waves are not affected by the electric boundary conditions; then, there are no stop bands and the waves in the ''gratings'' are the same as the plate modes discussed in Sec. II. The results reveal the important meaning of the ECC in the study on acoustic waves in metal gratings.

In the numerical calculation, the material parameters of PZT4 are taken from Ref. 14, and that of ZnO is taken from Ref. 18: $c_{44}=42.47(10^9\ \mathrm{N/m^3})$, $e_{15}=0.48$(C/m), $\varepsilon_{11}/\varepsilon_0=8.55$, $\rho=5680(\mathrm{kg/m^3})$ for ZnO; $c_{44}=30.00(10^9\ \mathrm{N/m^3})$, $e_{15}=10.5$(C/m), $\varepsilon_{11}/\varepsilon_0=804.6$, $\rho=7500(\mathrm{kg/m^3})$ for PZT4; $c'_{44}=116.0(10^9\ \mathrm{N/m^3})$, $\rho'=7800(\mathrm{kg/m^3})$ for steel.

IV. CONCLUSIONS

A theoretical analysis of the electromechanical coupling coefficient of SH acoustic waves in piezoelectric bounded plate is presented. This is important for acoustic wave device designs; for example, it helps to select a suitable frequency and structure parameters to excite a specific mode. SH acoustic wave propagation in piezoelectric bounded plates with metal gratings is analyzed as well. The nature and spectral characteristics of the stop bands are investigated. The results reveal the importance of the ECC in the study on acoustic waves in metal gratings.

ACKNOWLEDGMENT

This paper is supported by the Natural Scientific Foundation of Xi'an Jiaotong University.

[1] H. Kitabayashi and P. M. Smith, ''Analysis of SAW propagation in gratings on ZnO/Diamond substrates,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, No. 1, 249–261 (2001).
[2] T. Sato and H. Abe, ''Propagation of longitudinal leaky surface waves under periodic metal grating structure on lithium tetraborate,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, No. 2, 394–408 (1998).
[3] M. Hofer and R. Lerch, ''Finite element calculation of wave propagation and excitation in periodic piezoelectric systems,'' Fifth World Congress on Computational Mechanics 7–12 July, Vienna, Austria (2002).
[4] V. P. Plessky and T. Thorvaldsson, ''Periodic Green's functions analysis of SAW and leaky SAW propagation in a periodic system of electrodes on a piezoelectric crystal,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, No. 2, 280–293 (1995).
[5] K. Bløtekjar, K. A. Ingebrigtsen, and H. Skeie, ''A method for analysis wave in structures consisting of metal strips on dispersive media,'' IEEE Trans. Electron Devices **ED-20**, No. 12, 1133–1138 (1973).
[6] H. F. Tiersten, ''Wave propagation in an infinite piezoelectric plate,'' J. Acoust. Soc. Am. **35**, 53–58 (1963).
[7] J. L. Bleustein, ''Some simple modes of wave propagation in an infinite piezoelectric plates,'' J. Acoust. Soc. Am. **45**, 614–620 (1969).
[8] M. Romeo, ''Electromagneto-acoustic surface waves on dispersive piezoelectric layered media,'' J. Acoust. Soc. Am. **116**, 1488–1497 (2004).
[9] F. Simonetti, ''Lamb wave propagation in elastic plates coated with viscoelastic materials,'' J. Acoust. Soc. Am. **115**, 2041–2053 (2004).
[10] M. Castaings and B. Hosten, ''Guided waves propagating in sandwich structures made of anisotropic, viscoelastic, composite materials,'' J. Acoust. Soc. Am. **113**, 2622–2634 (2003).
[11] V. K. Tewary, ''Theory of elastic wave propagation in anisotropic film on anisotropic substrate: TiN film on single-crystal Si,'' J. Acoust. Soc. Am. **112**, 925–935 (2002).
[12] A. N. Darinskii, S. V. Biryukov, and M. Weihnacht, ''Fundamental frequency degeneracy of standing surface acoustic waves under metallic gratings on piezoelectric substrates,'' J. Acoust. Soc. Am. **112**, 2003–2013 (2002).
[13] Q. Wang and V. K. Varadan, ''Wave propagation in piezoelectric bounded plates by use of interdigital transducer. II. Wave excitation by interdigital transducer,'' Int. J. Solids Struct. **39**, 1131–1144 (2002).
[14] Q. Wang and V. K. Varadan, ''Wave propagation in piezoelectric bounded plates by use of interdigital transducer. I. Dispersion characteristics,'' Int. J. Solids Struct. **39**, 1119–1130 (2002).
[15] T. Pastureaud, V. Laude, and S. Ballandras, ''Stable scattering-matrix method for surface acoustic waves in piezoelectric multilayers,'' Appl. Phys. Lett. **80**, No. 14, 2544–2546 (2002).
[16] M. J. S. Lowe, ''Matrix techniques for modeling ultrasonic waves in multilayered media,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, No. 4, 525–542 (1995).
[17] K.-Y. Hashimoto, *Surface Acoustic Wave Devices in Telecommunications Modeling and Simulation* (Springer Berlin, 2000).
[18] D. F. Crisler, J. J. Cupal, and A. R. Moore, ''Dielectric, piezoelectric, and electromechanical coupling constants of zinc oxide crystal,'' Proc. IEEE **56**, 225–226 (1968).

Fast calculation of pulsed photoacoustic fields in fluids using *k*-space methods

B. T. Cox[a] and P. C. Beard
Department of Medical Physics and Bioengineering, University College London, Gower Street, London WC1E 6BT, United Kingdom

(Received 8 December 2004; revised 31 March 2005; accepted 31 March 2005)

Two related numerical models that calculate the time-dependent pressure field radiated by an arbitrary photoacoustic source in a fluid, such as that generated by the absorption of a short laser pulse, are presented. Frequency-wavenumber (*k*-space) implementations have been used to produce fast and accurate predictions. Model I calculates the field everywhere at any instant of time, and is useful for visualizing the three-dimensional evolution of the wave field. Model II calculates pressure time series for points on a straight line or plane and is therefore useful for simulating array measurements. By mapping the vertical wavenumber spectrum directly to frequency, this model can calculate time series up to 50 times faster than current numerical models of photoacoustic propagation. As the propagating and evanescent parts of the field are calculated separately, model II can be used to calculate far- and near-field radiation patterns. Also, it can readily be adapted to calculate the velocity potential and thus particle velocity and acoustic intensity vectors. Both models exploit the efficiency of the fast Fourier transform, and can include the frequency-dependent directional response of an acoustic detector straightforwardly. The models were verified by comparison with a known analytic solution and a slower, but well-understood, numerical model. [DOI: 10.1121/1.1920227]

PACS numbers: 43.35.Ud, 43.20.Px [YHB] Pages: 3616–3627

I. INTRODUCTION

There are several mechanisms by which electromagnetic radiation incident on a solid or fluid can generate acoustic waves, e.g., ablation, electrostriction, and thermoelastic expansion.[1] This paper is concerned with the prediction of the acoustic field generated in a fluid due to thermoelastic expansion following localized absorption of an electromagnetic pulse.

A number of papers present analytical solutions for the photoacoustic pressure under specific circumstances. In particular, Tam[1] quotes analytical solutions for the pressure for narrow and Gaussian laser beams in weakly absorbing fluids. Diebold *et al.* have presented many analytical solutions for the acoustic pressure caused by a variety of geometries of photoacoustic sources, including among others infinitely long Gaussian sources, point sources, and solid spheres.[2–6] All of these analytical solutions have the disadvantage that they are restricted to a particular case; the photoacoustic source must have a certain geometry. A time-domain numerical model of photoacoustic propagation that can include an arbitrary initial pressure distribution has been described in the literature[7–9] and is compared, in Sec. VI A, to the *k*-space models derived here. It is based on Poisson's integral solution to the wave equation[10] and provides an intuitive method useful for predicting the time evolution of the pressure at a point. For applications that require the pressure to be known at many points and small distances from the source it is considerably slower than the *k*-space models.[9]

Two models—model I and model II—that calculate the field generated by an arbitrarily shaped initial pressure distribution (or photoacoustic source) are presented in this paper. These propagation models fall into the category of wavenumber integration algorithms. Propagation models based on numerically solving a wavenumber integral are in widespread use in underwater acoustics and seismology.[11] This is mainly because these techniques lend themselves to the use of the fast Fourier transform (FFT) algorithm, and are therefore computationally efficient. Similar techniques have not previously been applied to the specific problem of the propagation of photoacoustic signals.

Model I uses an exact time propagator to calculate the acoustic field at all points on a grid for a single time following the absorption of an electromagnetic pulse. Unlike finite difference methods, in which the time step must be small to avoid instability, the acoustic field at any time may be predicted in one step without the need to calculate the field at intermediate times. With this model, the evolution of 3D fields through time can be visualized.

Model II, rather than calculating the pressure everywhere at a single time, calculates the pressure on a chosen line or plane for many times at once. It maps the pressure as a function of vertical spatial wavenumber to the pressure as a function of temporal frequency which results in a significant increase in speed over model I. For applications that do not require knowledge of the entire spatial field, e.g., for simulating the temporal signals detected by an array of sensors, this may be useful. The propagating and evanescent parts of the field are calculated separately in this model. As it is often convenient to neglect the evanescent part of the field, this model can show under which circumstances this is reasonable. In addition, model II may be used to generate near- and

[a]Electronic mail: bencox@medphys.ucl.ac.uk

 0001-4966/2005/117(6)/3616/12/$22.50

far-field single-frequency or broadband radiation patterns of photoacoustic sources, and simply adapted to calculate velocity potential, and thus particle velocity and acoustic intensity vectors.

These two *k*-space models allow photoacoustic fields to be calculated as much as 50 times faster than previous numerical techniques, with high accuracy as determined by comparisons with analytical results. They are more general than known analytical solutions, which usually include restrictive assumptions about the photoacoustic source geometry, and more efficient than current time-domain models. Both models can include the effect of an arbitrary, complex, frequency-dependent, directional detector response on the measured pressure, thus simulating not just the photoacoustic wave propagation but also its measurement. This is crucial for accurate simulations of measurements made by an array of detectors, for instance, when studying acoustic inverse problems such as photoacoustic imaging.[12–15] Indeed, photoacoustic imaging is one of the growing number of applications of photoacoustics to which these propagation models are applicable.

II. PHOTOACOUSTIC WAVE EQUATION

If a region of a fluid is heated, through the absorption of a laser pulse, a sound wave will be generated. Consider a stationary fluid with isotropic acoustic properties. Under conditions whereby the sound generation mechanism is thermoelastic, and terms containing the viscosity and thermal conductivity are negligible (thermal confinement),[16] the acoustic pressure, in the linear approximation, obeys the wave equation

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{-\beta}{C_p}\frac{\partial \mathcal{H}}{\partial t}, \tag{1}$$

where c is the sound speed, β is the volume thermal expansivity, C_p is the specific heat capacity, and $\mathcal{H}$ is the heat energy per unit volume and per unit time deposited in the fluid; p and $\mathcal{H}$ will depend, in general, on the position $\mathbf{x}=(x,y,z)$ and time t.

The heating is caused by the absorption of light. If the light fluence rate at a point in the fluid is $F(\mathbf{x},t)$ and the absorption distribution $\mu_a(\mathbf{x})$, then the heating function can be written as $\mathcal{H}(\mathbf{x},t)=\mu_a(\mathbf{x})F(\mathbf{x},t)$. Note that the fluence rate F will in general depend on the absorption $\mu_a(\mathbf{x})$ and scattering; hence, the heating function is nonlinearly related to the absorption distribution.

A. Instantaneous heating

When the laser pulse is short, so the density of the fluid has no time to change, it may be modeled as a δ function. This is satisfied if the duration of the laser pulse t_p is much shorter than the time it takes sound to travel across the heated region, a condition known as stress confinement.[16] In this case, the heating function can be written as

$$\mathcal{H}(\mathbf{x},t)=H(\mathbf{x})\,\delta(t), \tag{2}$$

where $H(\mathbf{x})$ is the heat deposited in the fluid per unit volume.[17] In this idealized case, all of the optical energy will be absorbed before the fluid density has time to change. Under this isochoric condition, the increase in the temperature T' of the heated fluid region is related to the absorbed energy H by C_v, the constant volume specific heat capacity, and ρ, the ambient density: $T'=H/(\rho C_v)$. The thermodynamic relation $\rho'=\rho\kappa_T p'-\beta\rho T'$, where ρ' and p' are small changes in density and pressure, respectively, holds true for constant isothermal compressibility κ_T and volume thermal expansivity β. If there is no change in the density, so $\rho'=0$, then $p'=(\beta/\kappa_T)T'$. Using the expression for T' above and noting that $\kappa_T=\gamma/\rho c^2$, where γ is the specific heat ratio, the increase in pressure due to the absorption of the laser pulse, from here on called the initial pressure distribution $p_0(\mathbf{x})$, may be written

$$p_0(\mathbf{x})=\left(\frac{\beta c^2}{C_p}\right)H(\mathbf{x})=\Gamma H(\mathbf{x}), \tag{3}$$

where Γ is the Grüneisen coefficient, a dimensionless constant that represents the efficiency of the conversion of heat to pressure. For water at room temperature $\Gamma\approx 0.11$.

With the heating function given by Eq. (2), the solution to Eq. (1), in the absence of acoustic boundaries, can be written in terms of the Green's function as

$$p(\mathbf{x},t)=\frac{\beta}{C_p}\int_0^\infty\int_V G(\mathbf{x},t;\mathbf{x}',t')H(\mathbf{x}')\,\delta'(t')\,\mathrm{d}\mathbf{x}'\,\mathrm{d}t', \tag{4}$$

where δ' represents the derivative of the delta function, and the free-space Green's function G is a solution to

$$\nabla^2 G-\frac{1}{c^2}\frac{\partial^2 G}{\partial t^2}=-\delta(\mathbf{x}-\mathbf{x}')\,\delta(t-t'), \tag{5}$$

and is often written as

$$G(\mathbf{x},t;\mathbf{x}',t')=\frac{\delta[|\mathbf{x}-\mathbf{x}'|-c(t-t')]}{4\pi|\mathbf{x}-\mathbf{x}'|}, \tag{6}$$

representing a spherical wave traveling outward from $\mathbf{x}'$. Using the property of the δ function deriviative $\int\delta'(t-t_0)f(t)\mathrm{d}t=-f'(t_0)$, and noting that $\delta G/\partial t'=-\partial G/\partial t$, Eq. (4) becomes

$$p(\mathbf{x},t)=\frac{\beta}{C_p}\int_V H(\mathbf{x}')\frac{\partial G}{\partial t}(\mathbf{x},t;\mathbf{x}',t')\mathrm{d}\mathbf{x}'. \tag{7}$$

With Eq. (3), this gives the pressure at time t as

$$p(\mathbf{x},t)=\frac{1}{c^2}\int_V p_0(\mathbf{x}')\frac{\partial G}{\partial t}(\mathbf{x},t;\mathbf{x}',t')\mathrm{d}\mathbf{x}'. \tag{8}$$

B. Initial value problem

It is instructive to see that, whereas in the previous section the wave equation included a source term, in the case of instantaneous heating this problem can be recast as an initial value problem with no explicit source term but with the distribution of pressure at the instant of the laser pulse, $p_0(\mathbf{x})$, taken as given. This makes the two initial conditions required for a unique solution explicit. We solve the homogeneous wave equation [Eq. (1) with no source term] with the two initial conditions

$$p|_{t=0}=p_0(\mathbf{x}), \quad \left.\frac{\partial p}{\partial t}\right|_{t=0}=0. \tag{9}$$

The first condition defines the acoustic pressure distribution at $t=0$, the instant of the pulse. The second initial condition is equivalent to assuming the particle velocity $\mathbf{v}(\mathbf{x})$ is initially zero everywhere. In the absence of acoustic boundaries the solution to this initial value problem may be written in terms of the Green's function, G, and the initial conditions as[18]

$$p(\mathbf{x},t)=\frac{1}{c^2}\int_V\left[G\frac{\partial p}{\partial t'}-p\frac{\partial G}{\partial t'}\right]_{t'=0}\mathrm{d}\mathbf{x}'. \tag{10}$$

Using Eq. (9), we see that this solution is identical to Eq. (8).

C. Green's function in *k* space

In this paper we are interested in k-space methods for calculating the pressure field. To find an expression for the Green's function in terms of frequency and wavenumbers, we take a 4D Fourier transform of Eq. (5) with respect to t and $\mathbf{x}$

$$-k^2G(\omega,\mathbf{k})+\frac{\omega^2}{c^2}G(\omega,\mathbf{k})=-\mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}'}\mathrm{e}^{i\omega t'}. \tag{11}$$

The free-space Green's function G may then be written in terms of the wavenumber vector $\mathbf{k}=(k_x,k_y,k_z)$ and frequency ω as a fourfold inverse Fourier transform

$$G(\mathbf{x},t;\mathbf{x}',t')=\frac{1}{(2\pi)^4}\int\int\frac{\mathrm{e}^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\mathrm{e}^{-i\omega(t-t')}}{k^2-(\omega/c)^2}\mathrm{d}\omega\,\mathrm{d}\mathbf{k}, \tag{12}$$

where $k=|\mathbf{k}|$. In Eq. (6) the Green's function is written as a spherical wave; in Eq. (12) this spherical wave is expressed as a sum of plane waves with direction given by $\mathbf{k}$ and frequency ω.

The two algorithms described below for calculating $p(\mathbf{x},t)$ from the initial pressure $p_0(\mathbf{x})$ are derived by analytically evaluating, in the first case, the ω integral in Eq. (12) and, in the second case, the integral over the vertical wavenumber k_z. The first results in a solution for the whole field at one instant in time, and the second in a method that calculates a pressure time series at all points on a line or plane. In both cases the singularity in the integrand in Eq. (12) at $\omega=ck$ is dealt with using contour integration.

III. MODEL I: THE WHOLE FIELD FOR ONE TIME

As noted above, the integrand in Eq. (12) is singular when $\omega=ck$. However, we can evaluate the integral using Cauchy's residue theorem. First, the difference of the two squared terms in the denominator is rewritten so the ω integral in Eq. (12), with t' set to zero, becomes

$$\int\frac{\mathrm{e}^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\mathrm{e}^{-i\omega t}}{(k-\omega/c)(k+\omega/c)}\mathrm{d}\omega, \tag{13}$$

from which it is clear that there are two simple poles on the real ω axis at $\pm ck$. This can be solved using Cauchy's residue theorem[18] to give Eq. (12), for $t>0$, as

$$G(\mathbf{x},t;\mathbf{x}')=\frac{c}{(2\pi)^3}\int\frac{\sin(ckt)}{k}\mathrm{e}^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\mathrm{d}\mathbf{k}. \tag{14}$$

The Green's function is now a sum of plane waves with different spatial frequencies multiplied by a time propagator. The time derivative of G is

$$\frac{\partial G}{\partial t}=\frac{c^2}{(2\pi)^3}\int\cos(ckt)\mathrm{e}^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\mathrm{d}\mathbf{k}. \tag{15}$$

Substituting this into Eq. (8) gives a solution for the pressure in a free-field given an initial pressure distribution

$$p(\mathbf{x},t)=\frac{1}{(2\pi)^3}\int\int p_0(\mathbf{x}')\cos(ckt)\mathrm{e}^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\mathrm{d}\mathbf{k}\,\mathrm{d}\mathbf{x}'. \tag{16}$$

Changing the order of the integration gives a two-stage method for calculating $p(\mathbf{x},t)$ at a given time t. First, the 3D spatial Fourier transform of the initial pressure distribution is taken (here the primes have been dropped)

$$p_0(\mathbf{k})=\int p_0(\mathbf{x})\mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}\mathbf{x}, \tag{17}$$

and second, the pressure at time t is calculated using

$$p(\mathbf{x},t)=\frac{1}{(2\pi)^3}\int p_0(\mathbf{k})\cos(ckt)\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}\mathbf{k}. \tag{18}$$

So, if the heating function $H(\mathbf{x})$ is known, $p_0(\mathbf{x})$ is known from Eq. (3), and the acoustic pressure at all positions and subsequent times can be calculated using Eqs. (17) and (18). To calculate the field at any time t thus requires just two 3D FFTs and one multiplication. Because the changes of p over time are calculated using the exact propagator $\cos(ckt)$ and not from an approximation, it is not necessary to calculate the field at intermediate times, as it is with finite difference methods, for instance. In practice the pressure is calculated on a grid of points, in which case the grid spacing must meet the usual Nyquist criterion to avoid aliasing in the spatial domain; it must be less than half the minimum wavelength. This first k-space method is similar to that proposed by Healey *et al.*[19] As the field is calculated everywhere in one step, it is a useful method for visualizing the field from a source at a particular time. Illustrations of this are given in Sec. VII A.

IV. MODEL II: TIME SERIES

Model I calculates the spatial distribution of the field everywhere for one instant of time. This can be slow if the pressure at just a single point or a few points is required as a function of time. Model II calculates a pressure time series at points on a line or plane. For applications in which a time series is required at only a few points in the field, or for simulating the signals measured by an array of detectors, this method may be much faster than model I.

In the above section the ω integral in Eq. (12) was calculated analytically. This resulted in a 3D Fourier transform in spatial wavenumbers to calculate the pressure field at one time. To arrive at a method that calculates the pressure at many times (a time series) in a single step, we need to leave

the Fourier transform with respect to ω in the expression. To this end we consider analytically evaluating the integral over k_z, the vertical component of the wavenumber vector $\mathbf{k}$, as a means of removing the singularities in Eq. (12). The simple poles at $k_z=\pm\zeta$ are clear if we rewrite this integral as

$$\int \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}e^{-i\omega t}}{(k_z-\zeta)(k_z+\zeta)}\,dk_z, \tag{19}$$

where we have chosen

$$\begin{aligned}\zeta&=\mathrm{sgn}(\omega)\sqrt{(\omega/c)^2-k_r^2} \quad \text{for } |\omega/c|\geqslant k_r,\\ \zeta&=+i\sqrt{k_r^2-(\omega/c)^2} \quad \text{for } |\omega/c|<k_r.\end{aligned} \tag{20}$$

Here, $k_r=\sqrt{k_x^2+k_y^2}$ and $\mathrm{sgn}(\omega)=+1$ for $\omega\geqslant 0$ and -1 for $\omega<0$. As before, Eq. (19) can be solved using Cauchy's residue theorem[20] to give the Green's function for $t>0$ as

$$G(\mathbf{x},t;\mathbf{x}')=\frac{i\pi}{(2\pi)^4}\int\int\int\left(\frac{1}{\zeta}\right)\times e^{i[k_x(x-x')+k_y(y-y')+\zeta|z-z'|-\omega t]}\,d\omega\,dk_x\,dk_y. \tag{21}$$

Compare the Green's function here, written as a sum of plane waves in x, y, and t, with the formulation in Eq. (12) as a sum of plane waves in $\mathbf{x}$. The time derivative of G is

$$\frac{\partial G}{\partial t}=\frac{\pi}{(2\pi)^4}\int\int\int\left(\frac{\omega}{\zeta}\right)\times e^{i[k_x(x-x')+k_y(y-y')+\zeta|z-z'|-\omega t]}\,d\omega\,dk_x\,dk_y, \tag{22}$$

and, using Eq. (8), the pressure $p(\mathbf{x},t)$ may be written

$$p(\mathbf{x},t)=\frac{\pi}{(2\pi)^4c^2}\int_V\int\int\int\left(\frac{\omega}{\zeta}\right)p_0(\mathbf{x}')\times e^{i[k_x(x-x')+k_y(y-y')+\zeta|z-z'|-\omega t]}\,d\omega\,dk_x\,dk_y\,d\mathbf{x}'. \tag{23}$$

Because of the two definitions of ζ, Eq. (20), it is convenient to consider the pressure field as a sum of propagating and evanescent components

$$p(\mathbf{x},t)=p_{\mathrm{prop}}(\mathbf{x},t)+p_{\mathrm{evan}}(\mathbf{x},t). \tag{24}$$

These are considered separately below.

A. Propagating (radiating) part of the field

In Eq. (23) ζ is really just shorthand for $\mathrm{sgn}(\omega)\sqrt{(\omega/c)^2-k_r^2}$ or $+i\sqrt{k_r^2-(\omega/c)^2}$, depending on whether $|\omega/c|\geqslant k_r$ or $<k_r$ [see Eqs. (20)]. ζ is real when $|\omega/c|\geqslant k_r$. This part of the solution, p_{prop}, consists of plane waves that propagate away from the source region and eventually form the acoustic far field. For this propagating, or radiating, part of the acoustic field, the integral over ω in Eq. (23) may be rearranged to give

$$\begin{aligned}&\int_{|\omega|\geqslant ck_r}\left|\frac{\omega}{\zeta}\right|e^{i\zeta|z-z'|}e^{-i\omega t}\,d\omega\\ &=\int_{ck_r}^{\infty}\left(\frac{\omega}{\zeta}\right)\{e^{i\zeta|z-z'|}e^{-i\omega t}+e^{-i\zeta|z-z'|}e^{i\omega t}\}d\omega\\ &=2\Re\left\{\int_{ck_r}^{\infty}\left(\frac{\omega}{\zeta}\right)e^{i\zeta|z-z'|}e^{-i\omega t}\,d\omega\right\},\end{aligned} \tag{25}$$

where $\Re$ indicates the real part. Equation (23) calculates the pressure on the z plane. We can set $z=0$ without loss of generality. In this case the propagating part of the field may be written as

$$p_{\mathrm{prop}}(x,y,t)=\frac{1}{(2\pi)^3c^2}\Re\left\{\int\int\int\left(\frac{\omega}{\zeta}\right)p_0(k_x,k_y,\omega)\times e^{i(k_xx+k_yy-\omega t)}\,dk_x\,dk_y\,d\omega\right\}, \tag{26}$$

where the ω integral is from ck_r to ∞ and $p_0(k_x,k_y,\omega)$ is interpolated from $p_0(k_x,k_y,\zeta)$, below, using Eq. (20).

$$p_0(k_x,k_y,\zeta)=\int\int\int p_0(\mathbf{x}')e^{-i(k_xx'+k_yy'-\zeta|z'|)}\,d\mathbf{x}'. \tag{27}$$

This interpolation from ζ to ω, vertical wavenumber to temporal frequency, effectively maps depth information in $p_0(z)$ to time information in $p(t)$. Indeed, in the 1D case where p_0 varies only with z, the temporal signal $p(t)$ has the same shape as the depth function $p(t)\propto p_0(t=z/c)$.[4] The complex exponential Fourier kernels in Eqs. (26) and (27) mean that most of the computations can be performed efficiently using FFTs.

In Sec. IV B we see that the evanescent part of the field can be written as a sine transform in ω. It is interesting that the propagating part, Eq. (26), can be written as a cosine transform in ω if

$$\Im\left\{\int\int p_0(k_x,k_y,\omega)e^{i(k_xx+k_yy)}\,dk_x\,dk_y\right\}=0, \tag{28}$$

where $\Im$ indicates the imaginary part. This condition requires both $\Im\{p_0(k_x,k_y,\zeta)\}$ to be odd in k_x and k_y and independent of ζ, and $\Re\{p_0(k_x,k_y,\zeta)\}$ to be even in k_x, k_y. Nothing, so far, has been assumed about $p(t)$ for $t<0$. As we are only interested in $p(t)$ for $t\geqslant 0$ we can choose $p(t)$ for $t<0$ to be any function. We choose to make $p(t)$ even, $p(t)=p(-t)$. As pressure is a real quantity, if $p(t)$ is even then its Fourier transform must be a real, even, function of frequency ω. As $p_0(\mathbf{x}')$ is real this can be achieved by removing the imaginary part of $\exp(i\zeta|z|)$ in Eq. (27) to get

$$p_0(k_x,k_y,\zeta)=\int\int\int p_0(\mathbf{x}')e^{-i(k_xx'+k_yy')}\cos(\zeta|z'|)\,d\mathbf{x}'. \tag{29}$$

If $p_0(\mathbf{x}')$ is symmetrical about $z'=0$ this may be written as a Fourier transform, like Eq. (17), except with $\mathbf{k}$ replaced with (k_x,k_y,ζ). [If $p_0(\mathbf{x}')$ is not symmetrical about $z'=0$ we can make it so, by using $1/2[p_0(x',y',z')+p_0(x',y',-z')]$ as the initial pressure. This will have no effect on the

pressure predicted on $z=0$ and does not require that $p_0(\mathbf{x}')=0$ for $z<0$, i.e., the initial pressure distribution is not restricted to one side of the measurement surface.]

The condition in Eq. (28) is now satisfied and Eq. (26) becomes a cosine transform

$$p_{\text{prop}}(x,y,t)=\frac{1}{(2\pi)^3 c^2}\int\int\int\left(\frac{\omega}{\zeta}\right)p_0(k_x,k_y,\omega)\times e^{i(k_x x+k_y y)}\cos(\omega t)\,dk_x\,dk_y\,d\omega. \quad (30)$$

An advantage of writing the propagating part as a cosine transform is that by inverting the transforms we arrive at a method of estimating $p_0(\mathbf{x})$ from measurements of $p(x,y,t)$ on the $z=0$ plane. This photoacoustic imaging method has been described by Köstli *et al.*[21] It can be used to estimate the initial pressure distribution from time series measurements of the acoustic field along a line or plane. However, this method ignores the evanescent component. An advantage, therefore, of having a forward model that computes the propagating and evanescent parts separately is that we can study under what conditions the evanescent component of the field is indeed negligible and this imaging algorithm gives accurate results.

B. Evanescent part of the field

ζ is imaginary when $|\omega/c|<k_r$. This second part of the solution from Eq. (24), p_{evan}, consists of evanescent plane waves that decay exponentially with z. (They are sometimes called inhomogeneous waves as they decay with z but not with x and y.) These waves do not contribute to the field beyond a few wavelengths of the source. They must, however, be included when calculating the acoustic pressure close to the source region. In this case the integral over ω in Eq. (23) can be rearranged to

$$\int_{|\omega|<ck_r}\left(\frac{\omega}{\zeta}\right)e^{i\zeta|z-z'|}e^{-i\omega t}\,d\omega = -2i\int_0^{ck_r}\left(\frac{\omega}{\zeta}\right)e^{i\zeta|z-z'|}\sin(\omega t)\,d\omega. \quad (31)$$

If, as above, we set $z=0$, then the evanescent part of the field can be written

$$p_{\text{evan}}(x,y,t)=\frac{-i}{(2\pi)^3 c^2}\int\int\int\left(\frac{\omega}{\zeta}\right)p_0(k_x,k_y,\omega)\times e^{i(k_x x+k_y y)}\sin(\omega t)\,dk_x\,dk_y\,d\omega, \quad (32)$$

where the ω integral is from 0 to ck_r. By remembering that ζ is purely imaginary for evanescent waves, this may be written, for comparison with Eq. (26), as

$$p_{\text{evan}}(x,y,t)=\frac{1}{(2\pi)^3 c^2}\Re\left\{\int\int\int\left(\frac{\omega}{\zeta}\right)p_0(k_x,k_y,\omega)\times e^{i(k_x x+k_y y-\omega t)}\,dk_x\,dk_y\,d\omega\right\}. \quad (33)$$

[Note that in this equation the ω integral is taken from 0 to ck_r, whereas in Eq. (26) it is from ck_r to ∞.] As for the propagating part of the field, $p_0(k_x,k_y,\omega)$ is interpolated from $p_0(k_x,k_y,\zeta)$, which is given by Eq. (27). However, here $p_0(k_x,k_y,\zeta)$ is required at imaginary values of ζ from 0 to ik_r in order to interpolate to $p_0(k_x,k_y,\omega)$ for $0\leqslant\omega<ck_r$ using Eq. (20). This means that, because ζ is imaginary, Eq. (27) is no longer a Fourier transform in z', as it is for the propagating part, but instead contains an $\exp(-|\zeta z|)$ term.

$$p_0(k_x,k_y,\zeta)=\int\int\int p_0(\mathbf{x}')e^{-i(k_x x'+k_y y')}e^{-|\zeta z'|}\,d\mathbf{x}'. \quad (34)$$

As the decaying exponential term cannot be evaluated with an FFT, it is more time consuming to calculate the evanescent part and so it is an advantage in terms of speed when this part of the field can be neglected, e.g., in the far field.

C. Including a reflecting plane boundary

In some circumstances it is convenient to include a reflection from a plane boundary in the initial pressure distribution by using an image of $p_0(\mathbf{x})$. If we add a perfectly reflecting acoustic boundary at z_0, with reflection coefficient V, then the initial pressure becomes $p_0(x,y,z)+Vp_0(x,y,2z_0-z)$. Note there is a subtlety with the use of a reflecting boundary. The bounding material must be modeled as a fluid, and not a solid, in order to avoid a pole in V in k space corresponding to an interface wave which occurs when both shear and compressional waves are allowed. The singularity causes aliasing problems due to undersampling; see Sec. V B.

D. Including an arbitrary detector response

To predict the pressure measured by a sensor, rather than the pressure at a single point, it is necessary to take into account the averaging effect due to its finite size. A wavenumber model of a planar detector response, $\mathcal{D}$, may be included straightforwardly by multiplying $p_0(k_x,k_y,\omega)$ in Eq. (30) by the detector response $\mathcal{D}(k_x,k_y,\omega)$. [In a similar way, the detector response can be included in model I by multiplying $p_0(\mathbf{k})$ in Eq. (18) by $\mathcal{D}(\mathbf{k})$.] As a simple example, a detector that averages the pressure over a small circle of radius a in the (x,y) plane has a directional response $\mathcal{D}(k_x,k_y)\equiv\mathcal{D}(k_r)\propto J_1(k_r a)/(k_r a)$. For a (multilayered) planar transducer[22,23] $\mathcal{D}$ can often be calculated using a wavenumber model.[24]

If $\mathcal{D}$ or V are complex, then it is necessary to use Eq. (26) to calculate the propagating part of the pressure and not Eq. (30), which requires condition Eq. (28) to hold. As there is no disadvantage in using Eq. (26) over Eq. (30), this presents no limitation in practice.

E. Implementation

In practice, this propagation model will be implemented with distances and wavenumbers discretized. In this case care must be taken over the sampling of the oscillating, and possibly ''spiky,'' integrands. In addition, there is a singularity at $\zeta=0$ in Eqs. (30) and (32). Consider the numerical integration of these equations. The singularity means that

however finely it is sampled it will always be undersampled; the samples can never properly reproduce the infinite peak. As undersampling in the Fourier domain appears as aliasing in the space domain, calculating the pressure field with Eq. (30) using FFTs, just as it stands without treating the singularity, will give inaccurate results. If the peak in the integrand can be smoothed to prevent the undersampling without sacrificing accuracy, then it will be possible to keep the solution in the form of a 3D FFT. One possible way to achieve this might be to shift the path of one of the integrals, say the k_x integral, off the real k_x axis into the complex plane, so the integral is over a contour from $-\infty+i\epsilon<k_x<\infty+i\epsilon$ instead of $-\infty<k_x<\infty$, thus moving away from the pole.[11] This approach is the subject of current research. The approach taken here is to remove the singularity with a change of variables. This is discussed in detail for the cylindrically symmetric case of model II in Sec. V. The disadvantage of this method is that the integral is no longer in the form of a 3D Fourier transform, but a 2D FFT and a sum. The sampling issue is also considered, for the cylindrically symmetric model, in Sec. V B.

V. MODEL II FOR A CYLINDRICALLY SYMMETRIC HEATING FUNCTION

It may be practical to reduce the 3D model to two dimensions when implementing these models. If a heating function $H(x,y,z)$ (heat deposited per unit volume) varies much less in one dimension, y say, than in the other two, it may be reasonable to assume it is constant with respect to y and calculate the pressure as a function of just x and z: $p(x,z,t)$. This 2D case can be straightforwardly implemented using the equations for the 3D case by setting $y=0$.

In many cases of practical interest the heating function H, and hence the initial pressure distribution, is cylindrically symmetric, with no dependence on angle. The output of a laser, or fiber-coupled laser, for instance, often exhibits cylindrical symmetry. In this case two of the Fourier transforms become a Hankel transform over a radial coordinate. This section describes the implementation of model II in this case, discussing algebraic forms of the heating function and the sampling requirements for the numerical calculations.

If $H(\mathbf{x}')=H(r',z')$ we can write the Fourier transforms over x' and y' in Eq. (27) or (34) as a Hankel transform over $r'=\sqrt{x^2+y^2}$. Dropping the primes, we have, for the propagating part

$$p_0(k_r,\zeta)=(2\pi)\int_{-\infty}^{\infty}\int_0^{\infty}p_0(r,z)e^{i\zeta|z|}J_0(k_r r)r\,dr\,dz, \tag{35}$$

and, for the evanescent part

$$p_0(k_r,\zeta)=(2\pi)\int_{-\infty}^{\infty}\int_0^{\infty}p_0(r,z)e^{-|\zeta z|}J_0(k_r r)r\,dr\,dz. \tag{36}$$

As before, $p_0(k_r,\omega)$ is interpolated from $p_0(k_r,\zeta)$ using Eqs. (20). To include the effect of a detector response $\mathcal{D}(k_r,\omega)$, we use $p_0(k_r,\omega)\mathcal{D}(k_r,\omega)$ in place of p_0, as in Sec. IV D. The integrals over k_x and k_y in Eqs. (30) and (32) also become Hankel transforms in $k_r=\sqrt{k_x^2+k_y^2}$ and the singularity when $\zeta=0$ is removed by making a change of variable. Changing from k_r to ϕ such that $k_r=(\omega/c)\cos\phi$ gives

$$p_{\text{prop}}(r,\omega)=A\omega^2\int_0^{\pi/2}p_0(\phi,\omega)\times J_0(r\omega\cos\phi/c)\cos\phi\,d\phi, \tag{37}$$

and, for the evanescent part, $k_r=(\omega/c)\cosh\phi$ gives

$$p_{\text{evan}}(r,\omega)=-A\omega^2\int_0^{\infty}p_0(\phi,\omega)\times J_0(r\omega\cosh\phi/c)\cosh\phi\,d\phi, \tag{38}$$

where $A=1/(2\pi c^3)$. The real part of the one-sided Fourier transform of $p_{\text{prop}}(r,\omega)$ gives the time series $p_{\text{prop}}(r,z=0,t)$. (If p_{prop} is assumed symmetrical about $t=0$ then this reduces to a one-sided cosine transform.) A sine transform of the evanescent part $p_{\text{evan}}(r,\omega)$ gives $p_{\text{evan}}(r,z=0,t)$. This derivation is for a detector at $z=0$, although it may be straightforwardly generalized to other detector depths.

A. Cylindrically symmetric heating functions

In the most general case the heating function $H(r,z)$, and hence initial pressure $p_0(r,z)=\Gamma H(r,z)$, will be given as values on a grid. In this case numerical Fourier and Hankel transforms[25,26] may be used to calculate $p_0(k_r,\zeta)$ from Eqs. (35) and (36). It may be, however, that $H(r,z)$ can be approximated by an algebraic expression. In such cases it may be possible to calculate the Hankel and Fourier transforms of $p_0(r,z)$ analytically. For instance, H may be approximated by a polynomial and the Hankel transform and Fourier transforms calculated by hand or looked up in a table. This can result in a significant increase in speed and accuracy as it removes the need for numerical Hankel transforms and interpolations.

Consider the example of a short laser pulse with a known beam profile, incident on the surface of a nonscattering fluid absorber (Fig. 1). The heating function may be separated into z- and r-dependent parts

$$H(r,z)=\mu_a E_0\hat{H}_r(r)\hat{H}_z(z), \tag{39}$$

where μ_a is the optical absorption coefficient, E_0 is the energy contained in a single pulse, and $\hat{H}_z(z)$ and $\hat{H}_r(r)$ are appropriately normalized shape functions that give the z and r dependence of the initial pressure distribution. If there is an acoustic impedance mismatch at the surface of the fluid that will cause reflections, then an image source may be included in $\hat{H}_z$ and, in this way, the acoustic effects of the boundary may be accounted for, as discussed in Sec. IV C. Examples of algebraic heating functions that might be useful in this scenario are given below. The depth dependence and radial dependence (beam profile) are treated separately.

1. Algebraic depth functions

For a pure absorber a distance z_0 above the measurement plane, the part of the heating function that depends on depth will be, according to Beer's law[27]

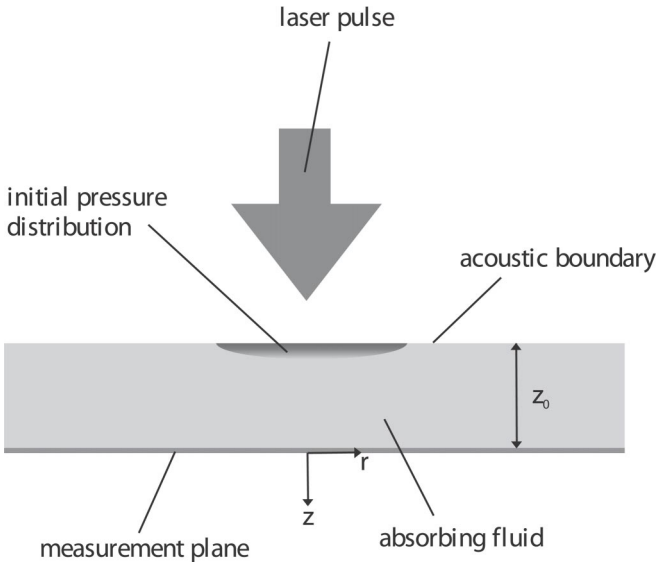


FIG. 1. An example of a cylindrically symmetric heating function, as described in Sec. V A.

$$\hat{H}_z(z) = \exp(-\mu_a(z+z_0)) \quad \text{for } z \geqslant -z_0$$
$$= 0 \quad \text{for } z < -z_0. \tag{40}$$

This assumes there is no acoustic impedance mismatch at the surface of the absorber. If, however, there is a transparent material with different acoustic properties above the absorber then there will be an acoustic reflection there. In general the reflection coefficient will depend on the horizontal wavenumber and frequency, $V(k_r, \omega)$. In this more general situation the depth function becomes

$$\hat{H}_z(z) = \exp(-\mu_a(z+z_0)) \quad \text{for } z \geqslant -z_0$$
$$= V \exp(\mu_a(z+z_0)) \quad \text{for } z < -z_0. \tag{41}$$

Substituting $\hat{H}_z(z)$ into the z integral (Fourier transform) in Eq. (35) we get

$$\hat{H}_z(\zeta) = \left(\frac{V e^{i\zeta z_0} + e^{-\mu_a z_0}}{\mu_a - i\zeta}\right) + \left(\frac{e^{i\zeta z_0} - e^{-\mu_a z_0}}{\mu_a + i\zeta}\right). \tag{42}$$

[If we use a cosine transform to transform the propagating part to the time domain, then we must use only the real part of Eq. (42). This is equivalent to calculating the cosine transform of $\hat{H}_z(z)$ using Eq. (29), instead of the full Fourier transform, and implicitly assumes $p(t) = p(-t)$.]

2. Algebraic beam profiles

For a laser beam with a Gaussian profile, the radial part of the heating function may be written as

$$\hat{H}_r(r) = e^{-(r/\sigma)^2}. \tag{43}$$

The Hankel transform of this is[28]

$$\hat{H}_r(k_r) = \left(\frac{\sigma^2}{2}\right) e^{-(k_r\sigma/2)^2}. \tag{44}$$

For a top hat beam profile:

$$\hat{H}_r(r) = 1 \quad \text{for } r \leqslant \sigma$$
$$0 \quad \text{for } r > \sigma, \tag{45}$$

which has a Hankel transform

$$\hat{H}_r(k_r) = \sigma \frac{J_1(k_r\sigma)}{k_r}. \tag{46}$$

B. Sampling

When Eqs. (37) and (38) are implemented the integrals will be approximated by summations. For this approximation to be accurate the oscillating integrands must be sampled more than once per oscillation. If the heating function is given by Eqs. (42) and (46), and if two adjacent samples of ϕ are separated by $\Delta\phi$, then the sampling criterion for Eq. (37) is

$$\Delta\phi < \frac{2\pi}{(r+z+\sigma)k}, \tag{47}$$

and for Eq. (38) is

$$\Delta\phi < \frac{2\pi}{k(r\sinh\phi_{\max} + z\cosh\phi_{\max} + \sigma\sinh\phi_{\max})}, \tag{48}$$

where $\phi_{\max}$ is the practical upper limit of the integral. If $\phi_{\max}$ is chosen, fairly arbitrarily, to be $\pi/2$, then this criterion is more stringent than that for the propagating part, as $\sinh(\pi/2) \approx 2.3$ and $\cosh(\pi/2) \approx 2.5$. This choice of $\phi_{\max}$ will be sufficiently high so long as $\hat{H}_r(k_r)$ is negligible above $k_r = (\omega/c)\cosh(\pi/2)$, which is about $2.5(\omega/c)$. From Eq. (44) we see that, when the heating function is a Gaussian, $\hat{H}_r(k_r)$ falls to 1% of its peak value when $(k_r\sigma)^2 = -4\ln(0.01)$ or $k_r = 4.3/\sigma$. So, if $4.3/\sigma < 2.5(\omega/c)$, or $\sigma > 1.7(c/\omega)$, then this sampling criterion will be sufficient. For a top hat source, Eq. (46) shows that $\hat{H}_r(k_r)$ is an oscillating function which falls to about 1% of its peak value when $k_r\sigma \approx 30$ and the criterion is $\sigma > 12(c/\omega)$.

As well as sampling sufficiently often to approximate accurately a rapidly oscillating integrand, it is necessary to sample sufficiently often to approximate spikes in the integrand. Undersampling a spike leads to aliasing or wraparound error in the transform domain. The wraparound error arising from $\hat{H}_z(\zeta)$ in Eq. (42) will be reduced by 60 dB if[11] the following conditions for the propagating and evanescent parts are met:

$$\Delta\phi < \frac{\mu\pi\log_{10}(e)}{3k}, \quad \Delta\phi < \frac{\mu\pi\log_{10}(e)}{3k\cosh\phi_{\max}}. \tag{49}$$

The ω integrals, Fourier transforms, that transform into the time domain must also be appropriately sampled. The criteria are that

$$\Delta\omega < \frac{\pi c}{r+z+\sigma}, \tag{50}$$

for the propagating part and

$$\Delta\omega < \frac{\pi c}{r\cosh\phi_{\max} + z\sinh\phi_{\max} + \sigma\cosh\phi_{\max}}, \tag{51}$$

for the evanescent part. The equivalent requirements to Eqs. (49) for the propagating and evanescent parts, respectively, are

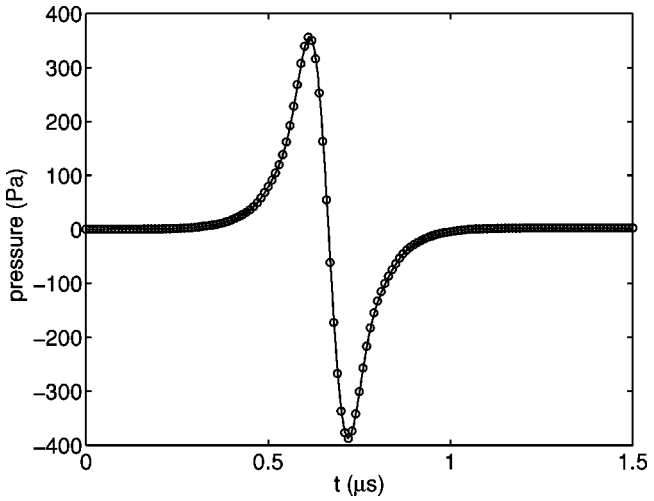


FIG. 2. Pressure 1 mm off axis and 1 mm below a pressure release boundary for an impulsive heating function with Gaussian radial profile. Model I (circles), model II (solid line). $\mu_a = 100\,\mathrm{cm}^{-1}$, $\sigma = 2$ mm, $V = -1$, $E_0 = 26\,\mu$J.

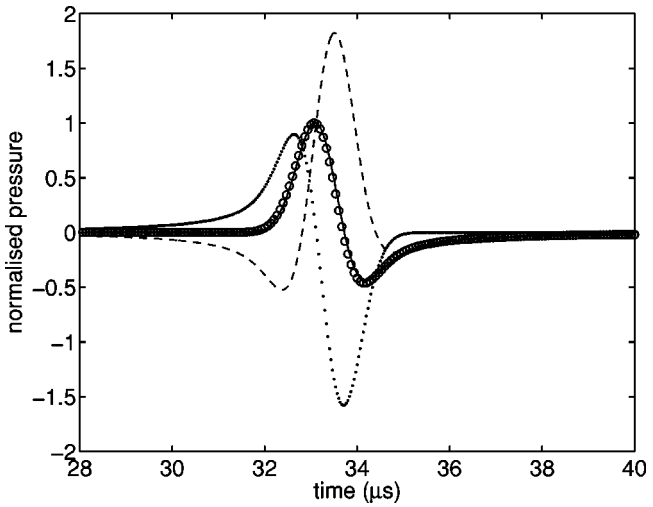


FIG. 3. Pressure 5 cm off axis for an infinitely long impulsive heating function with Gaussian radial profile. Diebold's analytic solution (circles), model II (solid) with the propagating (dashed), and evanescent (dotted) components also shown separately. $\mu_a = 0.01\,\mathrm{cm}^{-1}$, $\sigma = 1$ mm, $V = 1$.

$$\Delta\omega < \frac{\mu\pi c \log_{10}(\mathrm{e})}{3}, \quad \Delta\omega < \frac{\mu\pi c \log_{10}(\mathrm{e})}{3 \sinh \phi_{\max}}. \tag{52}$$

VI. TESTING THE MODELS

A. Time series calculations

Both models I and II can be used to calculate the time evolution of the pressure at a point, or at a receiver, due to an initial distribution $p_0(\mathbf{x})$. However, for signals containing high frequencies the mesh required for model I is large and the calculations are therefore slow. Model II, on the other hand, which calculates the pressure for many times at once, is better suited to time series calculation. Here, both models are compared to an analytical solution and to another model, described in the literature, that is based on a time-domain solution to the wave equation.

First, a simple test. Acoustic waves generated as described in Sec. II have the property that

$$\int_0^\infty p(t)\,\mathrm{d}t = 0. \tag{53}$$

This can be seen by considering the time integral of Eq. (8) with the Green's function given by Eq. (6). As the derivative of the delta function has the property $\int f(t)\delta'(t-t_0)\mathrm{d}t = -f'(t_0)$, and as the initial pressure distribution $p_0(\mathbf{x})$ is not a function of t, the time integral is zero. The time series generated by these numerical models were found to satisfy this property down to the machine precision.

To show that models I and II give the same pressure waveforms, the following example is calculated using both the cylindrically symmetric model II and model I (128^3 grid). The pressure at a point 1 mm off axis and 1 mm below a pressure-release acoustic boundary is shown in Fig. 2 for a Gaussian source [Eq. (43) with $\sigma = 2$ mm and $r = 1$ mm, and Eq. (41) with $\mu_a = 100\,\mathrm{cm}^{-1}$, $z_0 = 1$ mm, and $V = -1$]. The total energy in the pulse E_0 was 26 μJ. The circles show the pressure calculated using the 3D model I, and the solid line model II. Frequencies up to 5 MHz were included in the calculations. Clearly there is a good correspondence. As it is time consuming to calculate time series using model I, the remaining times series in this section are calculated using the cylindrically symmetric implementation of model II.

Diebold and Sun[4] present a solution to Eq. (1) for an infinitely long, cylindrical source with a Gaussian radial (beam) profile. We approximate such a source by setting $\mu_a = 0.01\,\mathrm{cm}^{-1}$ and $V = 1$ in Eq. (41). Frequencies up to 1 MHz were included. A comparison between Diebold's solution and model II for a point 5 cm off the source axis is given in Fig. 3. Diebold's solution is shown with circles. Model II (solid line) and its propagating (dashed) and evanescent (dotted) components are also shown. It is clear that in this case both the propagating and evanescent parts are required to describe the field correctly. The slight disagreement between the two (solid curve and circles) around 35 μs was due to the approximation of the infinite source by one that decays slowly ($\mu_a > 0$). It was necessary to make μ_a slightly larger than zero in order to satisfy the sampling conditions.

Poisson's solution to the wave equation[10] can be derived from Eqs. (6) and (8) and calculates a solution by summing over spherical surfaces S of radius $|\mathbf{x}-\mathbf{x}'| = ct$

$$p(\mathbf{x},t) = \frac{1}{4\pi c}\frac{\partial}{\partial t}\int_S \frac{p_0(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}\,\mathrm{d}S. \tag{54}$$

This time-domain model implicitly includes both propagating and evanescent components but is slow as it is not designed to use FFTs. Frenz *et al.*[8] use this model to calculate the acoustic field generated by the absorption of a pulse of light from an optical fiber by assuming the light generates a top hat heating function given by $H(z)$ in Eq. (41), with $\mu_a = 900\,\mathrm{cm}^{-1}$, $V = 1$, and $H(r)$ given by Eq. (45), with $\sigma = 300$ μm. (Actually they taper the edges of the beam profile slightly. We ignore this refinement for the sake of simplicity.) Time series were generated using these parameters with an implementation of the Poisson model described by Köstli and Beard[9] as well as with model II. The frequency response of both models was rolled off at high frequencies

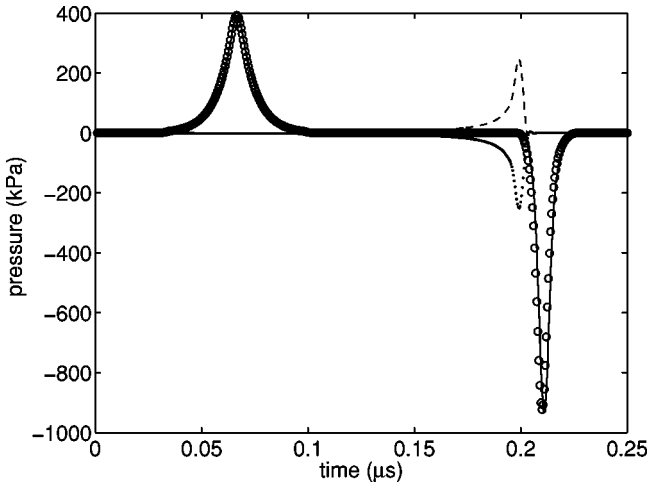


FIG. 4. Pressure on axis and 0.1 mm below the tip of an optic fiber (top hat profile) following impulsive heating calculated using Poisson model (circles) and model II (solid) with the propagating (dashed) and evanescent (dotted) components shown. $\mu_a = 900\ \mathrm{cm}^{-1}$, $\sigma = 300\ \mu\mathrm{m}$, $V = 1$, $E_0 = 26\ \mu\mathrm{J}$.

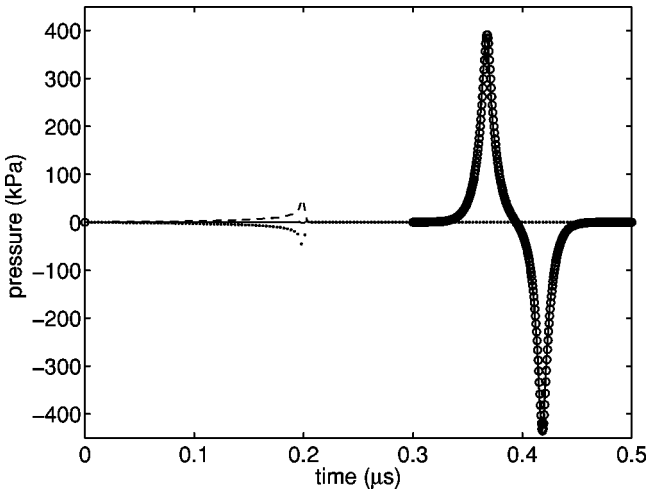


FIG. 6. Pressure on axis 0.55 mm below the tip of an optic fiber (top hat profile) following impulsive heating calculated using Poisson model (circles) and model II (solid) with the propagating (dashed) and evanescent (dotted) components shown. $\mu_a = 900\ \mathrm{cm}^{-1}$, $\sigma = 300\ \mu\mathrm{m}$, $V = 1$, $E_0 = 26\ \mu\mathrm{J}$.

(−3-dB point at 100 MHz). Figures 4–7 show the time series on axis at depths of 0.1 and 6.5 mm below the tip of the fiber, and at a depth of 0.55 mm both on axis and 0.3 mm off axis. The result from the Poisson model is shown as black circles, the total field from model II as a solid line, the propagating part of the field is a dashed line, and the evanescent part a dotted line. [These figures correspond to Figs. 3(a) and (b) in Frenz *et al.*]

In calculating the examples in Figs. 4 and 6, where the detector is close to the source region, model II (cylindrical version) was over 50 times faster than the Poisson model. For examples where the detector is further from the source region, though, there may be an advantage in using the Poisson model in that it can calculate the pressure for a range of times not starting from zero. Model II, on the other hand, always calculates the time series from $t = 0$. For the example in Fig. 5, to calculate the whole time series from zero to 4.4 μs the Poisson model would take 70 times as long as model II, but to calculate the portion of the time series shown took only three times as long. (This takes into account the fact that only the propagating waves contribute to the acoustic pressure in the far field, and it was therefore unnecessary to calculate the evanescent part, thereby decreasing the calculation time for model II at long distances.)

In the cylindrical version of model II used here, the singularity at $\zeta = 0$ was removed using a change of variables and the solution is thus no longer written as a Hankel transform from k_r to r. The time series for each radial position r must be calculated separately. However, if an alternative means of removing the singularity were used (see Sec. IV E), then the pressure time series at many values of r could be calculated at once via one fast Hankel transform (or via FFTs in the Cartesian case). The Poisson model does not have this po-

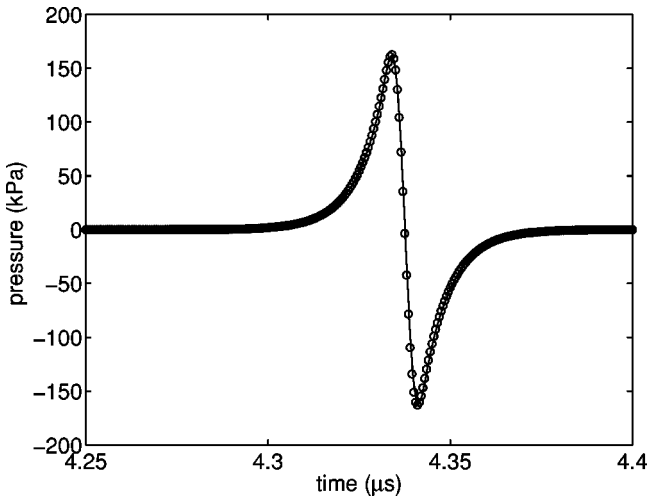


FIG. 5. Pressure on axis and 6.5 mm below the tip of an optic fiber (top hat profile) following impulsive heating calculated using Poisson model (circles) and model II (solid). The evanescent component is negligible. $\mu_a = 900\ \mathrm{cm}^{-1}$, $\sigma = 300\ \mu\mathrm{m}$, $V = 1$, $E_0 = 26\ \mu\mathrm{J}$.

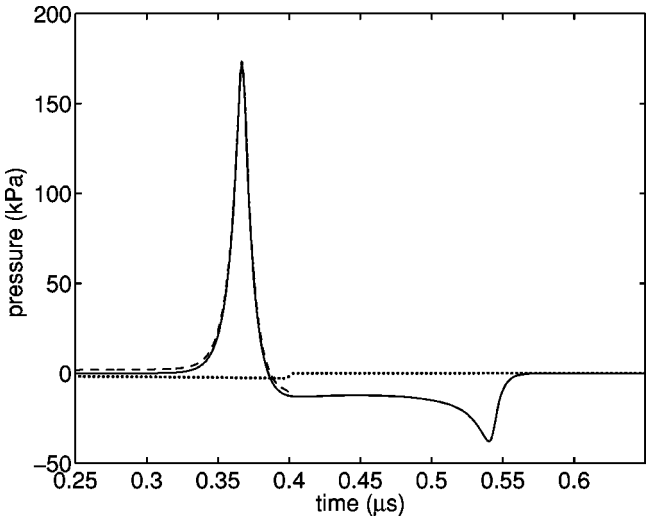


FIG. 7. Pressure 0.3 mm off axis and 0.55 mm below the tip of an optic fiber (top hat profile) following impulsive heating calculated using model II (solid) with the propagating (dashed) and evanescent (dotted) components shown. $\mu_a = 900\ \mathrm{cm}^{-1}$, $\sigma = 300\ \mu\mathrm{m}$, $V = 1$, $E_0 = 26\ \mu\mathrm{J}$.

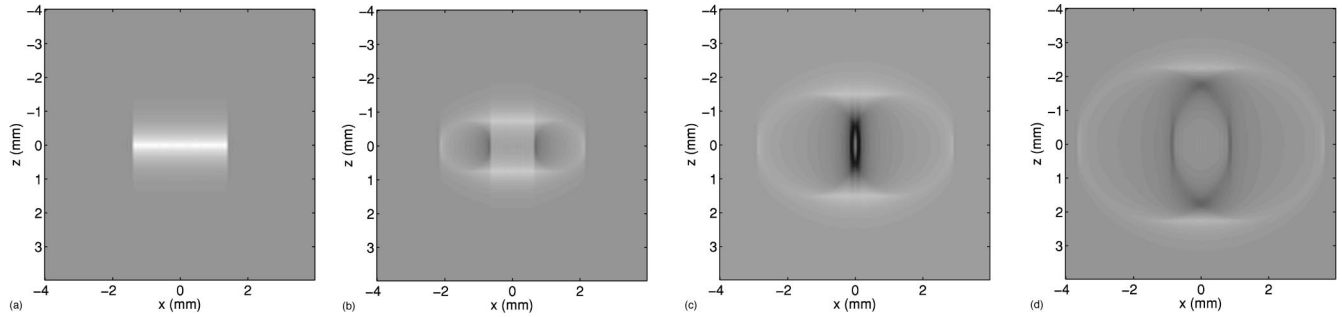


FIG. 8. 8×8 mm slices through an axis-symmetric 3D pressure field with a top hat initial pressure distribution. $\sigma=1.4$ mm, $\mu_a=25\ \mathrm{cm}^{-1}$, $V=1$ (reflecting boundary), $E_0=26\ \mu$J. From left to right, the four frames show the initial pressure distribution $p_0(\mathbf{x})$, then the pressure $p(\mathbf{x},t)$ at 0.5, 1, and 1.5 μs, respectively. The linear gray scale ranges from −1.6 kPa (black) to 1.2 kPa (white).

tential and is limited to calculating the pressure time series at one point. For the calculation of the signals arriving at a linear array of detectors, then, such an implementation of model II would be much more efficient for many practical situations.

B. Experimental validation

The comparisons above, between time series generated using the *k*-space models described here and both an analytical solution and a numerical time domain model, show that these *k*-space models correctly calculate solutions to Eq. (1). What they do not test are the assumptions made in deriving this equation (negligible heat conduction and viscosity, and linearity approximations). Experiments are required in order to test the applicability of Eq. (1) to photoacoustic generation in fluids. One challenge such experiments present is that, to ensure a fair comparison between model and experiment, it is necessary to know the frequency-dependent directional response of the ultrasound sensor, $\mathcal{D}$. This work is still progressing but initial experiments show good agreement between these models and such experiments.[29]

VII. OTHER APPLICATIONS OF THE MODELS

A. Model I: Visualization of the acoustic field

For calculating time series model I is slow, as the whole field is calculated at each instant of time. For visualizing the evolution of the whole field, though, it is ideally suited. The whole field can be viewed at any chosen time instant without the need to step forward in time, as is required in finite difference methods.

Figures 8–10 show examples of pressure fields calculated using model I. Each example shows 8×8-mm slices through an axis-symmetric 3D pressure field calculated on a 3D mesh of 128^3 points. (z and r both range from −4 to 4 mm.) In all of these examples the top half of the picture is there purely to enable calculation of the effect of a boundary on the acoustic field in the bottom half. The field that would be realized in practice, then, is just the bottom half. In Fig. 8 the initial pressure has a top hat profile in r and decays exponentially with $|z|$. In this case, the upper half of p_0 is an image source representing an acoustic reflection from a rigid boundary. An initial pressure distribution such as this, shown in the leftmost image, may be generated in practice by sending a laser pulse along a multimode fiber which terminates in a homogeneous optically absorbing liquid. The other three images are snapshots of the field at 0.5, 1.0, and 1.5 μs later. The negative pressure region that develops close to the rigid boundary—the dark area in the third picture—is due to the edge waves and shows clearly why cavitation has been observed in a case such as this. Indeed, Frenz *et al.* have photographed the cavitation and the edge waves generated in this way using Schlieren methods.[8,30]

Figure 9 is similar to Fig. 8 except that instead of modeling a solid–fluid interface at $z=0$ there is an acoustic pressure-release, air–fluid boundary. The reflection from this is represented by the negative (dark) region in the top half of the first image. This results in a large negative part to the bipolar signal, as the subsequent images show. Figure 10, in which the top hat profile is replaced with a Gaussian, shows that by removing the sharp radial discontinuity, the edge

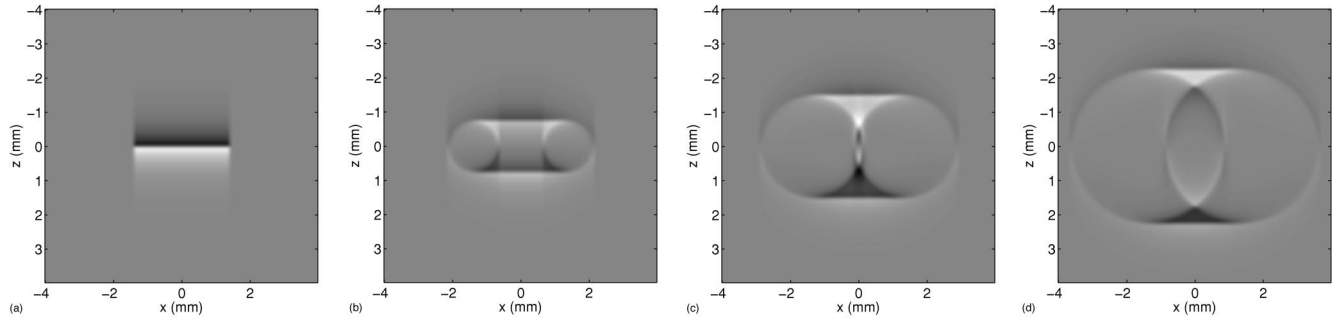


FIG. 9. Tophat laser pulse incident on an absorbing fluid at a pressure-release boundary. As Fig. 8 except $V=-1$. The frames show the pressure distribution $p(\mathbf{x},t)$ at $t=0$, 0.5, 1, and 1.5 μs, respectively. Gray scale range: −1.1 kPa (black) to 1.1 kPa (white).

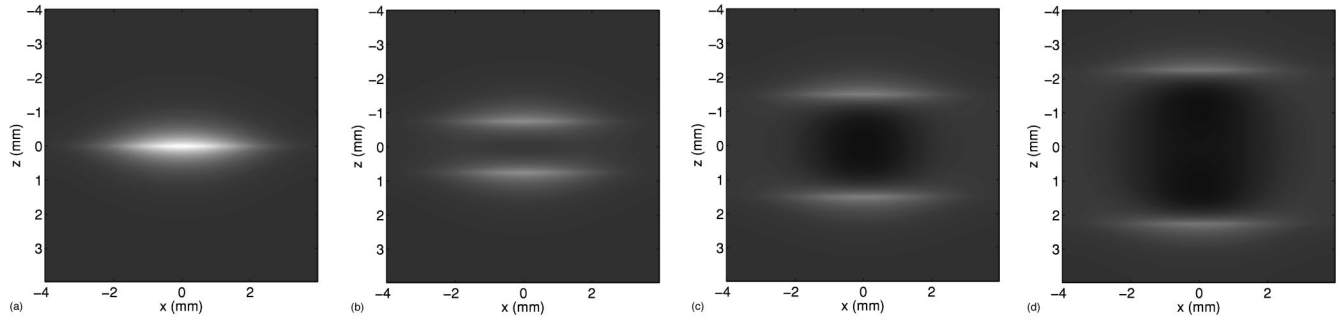


FIG. 10. As Fig. 8 except $H_r(r)$ has a Gaussian rather than a top hat profile (1/e point at $r=2$ mm). Edge waves are not generated. The frames show the pressure distribution $p(\mathbf{x},t)$ at $t=0$, 0.5, 1, and 1.5 μs, respectively. Gray scale range: -87 Pa (black) to 550 Pa (white).

waves are less distinct, and the field does not develop such a large negative pressure.

B. Model II: Simulation of array measurements

As we have seen above, model II can be used to generate photoacoustic time series efficiently. Figure 11 shows a simulation of time series measured along a linear array in response to photoacoustic point sources 1.5 and 3 mm above the detector plane. The time series are placed side by side and the linear gray scale indicates the acoustic pressure. (The white and black ends of the gray scale indicate positive and negative acoustic pressures, respectively.)

Acoustic array measurements are used in many applications including photoacoustic imaging,[21] in which the aim is to estimate the initial pressure distribution $p_0(\mathbf{x})$ from a set of time series measurements obtained over a line or plane. The models described here can be used to generate synthetic, noise-free, time series data in order to test aspects of such problems. For instance, the time series in Fig. 11 could represent the acoustic signals recorded by a line array of detectors perpendicular to two blood vessels which have been illuminated with a laser pulse. These data could be used to test photoacoustic image reconstruction algorithms.

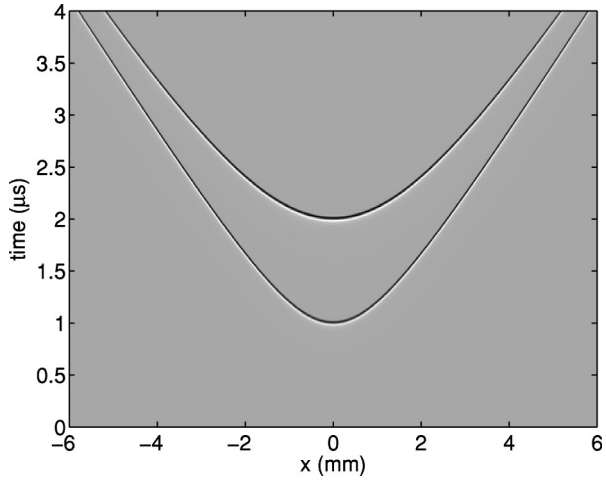


FIG. 11. Simulation of time series measured along a line array of detectors with two point photoacoustic sources at depths of 1.5 and 3 mm. The linear gray scale is in arbitrary units ranging from negative (black) to positive (white) pressures.

C. Model II: Radiation patterns

The emphasis in this paper is on the ability of model II to generate time series efficiently. However, as well as calculating time series, model II can be used to generate single-frequency or broadband radiation patterns of photoacoustic sources. The (complex) pressure at a single frequency may be calculated using Eqs. (26) and (32), without the final Fourier transform in time, or Eqs. (37) and (38). As above, the propagating and evanescent parts of the radiating field may be computed separately. Radiation patterns generated in this way are shown in Cox *et al.*[29]

In addition, because the velocity potential ψ is related to the complex pressure by $\psi=p/(i\omega\rho)$, model II can be simply adapted to calculate it. The complex particle velocity $\mathbf{v}$ and acoustic intensity $\mathbf{I}$ then follow using the standard relations: $\mathbf{v}=\nabla\psi$ and $\mathbf{I}=p\mathbf{v}$. In this way, model II can be used to calculate single-frequency and broadband intensity plots.

VIII. SUMMARY

Two related numerical models for calculating the acoustic field in a fluid following the absorption of a pulse of light have been described. Model I uses an exact time propagator to calculate the acoustic field at all points on a spatial grid for one moment of time in a single step. It can be used to visualize the 3D evolution of a photoacoustically generated wave field through time. Model II calculates pressure time series for points on a line or plane by mapping from the vertical spatial wavenumber to temporal frequency. The use of this mapping, and an FFT to perform part of the calculation, makes this an efficient model. As it calculates the propagating and evanescent parts of the field separately, model II can also be used to calculate the near- and far-field radiation patterns of photoacoustic sources and, with one small change, can be used to calculate velocity potential and thus particle velocity and acoustic intensity vectors. The effect on the measured pressure of a finite-sized planar detector with arbitrary frequency-dependent directional response can be included simply in either model—a necessary requirement when simulating measurements made with real detectors. Time series from both of these *k*-space models were tested against an analytical solution and a well understood but slower numerical model based on Poisson's solution to the wave equation.

ACKNOWLEDGMENTS

The authors are grateful to Kornel Köstli for programming the Poisson model used in Sec. VI A. This work has been supported by the Engineering and Physical Sciences Research Council, UK.

[1] A. C. Tam, "Applications of photoacoustic sensing techniques," Rev. Mod. Phys. **58**(2), 381–431 (1986).
[2] G. J. Diebold and P. J. Westervelt, "The photoacoustic effect generated by a spherical droplet in a fluid," J. Acoust. Soc. Am. **84**(6), 2245–2251 (1988).
[3] G. J. Diebold, T. Sun, and M. I. Khan, "Photoacoustic monopole radiation in 1-dimension, 2-dimension, and 3-dimension," Phys. Rev. Lett. **67**(24), 3384–3387 (1991).
[4] G. J. Diebold and T. Sun, "Properties of photoacoustic waves in one-dimension, 2-dimension, and 3-dimension," Acustica **80**(4), 339–351 (1994).
[5] M. I. Khan and G. J. Diebold, "The photoacoustic effect generated by an isotropic solid sphere," Ultrasonics **33**(4), 265–269 (1995).
[6] M. I. Khan and G. J. Diebold, "The photoacoustic effect generated by laser irradiation of an isotropic solid cylinder," Ultrasonics **34**(1), 19–24 (1996).
[7] G. Paltauf, H. SchmidtKloiber, and H. Guss, "Light distribution measurements in absorbing materials by optical detection of laser-induced stress waves," Appl. Phys. Lett. **69**(11), 1526–1528 (1996).
[8] M. Frenz, G. Paltauf, and H. SchmidtKloiber, "Laser-generated cavitation in absorbing liquid induced by acoustic diffraction," Phys. Rev. Lett. **76**(19), 3546–3549 (1996).
[9] K. Köstli and P. Beard, "Two-dimensional photoacoustic imaging by use of Fourier-transform image reconstruction and a detector with an anisotropic response," Appl. Opt. **42**(10), 1899–1908 (2003).
[10] L. Landau and E. Lifshitz, *Fluid Mechanics*, 2nd ed. (Butterworth Heinemann, Oxford, 1987).
[11] F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 2000).
[12] X. Wang, Y. Pang, G. Ku, X. Xie, G. Stoica, and L. V. Wang, "Noninvasive laser-induced photoacoustic tomography for structural and functional *in vivo* imaging of the brain," Nat. Biotechnol. **21**(7), 803–806 (2003).
[13] X. Wang, Y. Pang, G. Ku, G. Stoica, and L. V. Wang, "Three-dimensional laser-induced photoacoustic tomography of mouse brain with the skin and skull intact," Opt. Lett. **28**(19), 1739–1741 (2003).
[14] C. G. A. Hoelen, F. F. M. de Mul, R. Pongers, and A. Dekker, "Three-dimensional photoacoustic imaging of blood vessels in tissue," Opt. Lett. **23**(8), 648–650 (1998).
[15] P. C. Beard, "Photoacoustic imaging of blood vessel equivalent phantoms," Proc. SPIE **4618**, 54–62 (2002).
[16] G. Paltauf and P. E. Dyer, "Photomechanical processes and effects in ablation," Chem. Rev. (Washington, D.C.) **103**(2), 487–518 (2003).
[17] If the pulse is too long for stress confinement but short enough for thermal confinement then, if we can separate the heating function into $\mathcal{H}(\mathbf{x},t) = H(\mathbf{x})H_t(t)$, where $H_t(t)$ describes the temporal shape of the pulse, the pressure due to a longer pulse may be written as the convolution $\int p(t - t')H_t(t')\mathrm{d}t'$, where $p(t)$ is the pressure due to δ function heating. Note that a moving source is not separable into spatial and temporal parts, and so cannot be modeled using the methods described in this paper, unless the source is moving much more slowly than the speed of sound and may be approximated as stationary.
[18] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962), pp. 183–187.
[19] A. J. Healey, S. Leeman, and J. Weight, "Space-time imaging of transient ultrasound fields," Int. J. Imaging Syst. Technol. **8**, 45–51 (1997).
[20] M. S. Howe, *Acoustics of Fluid-Structure Interactions* (Cambridge University Press, Cambridge, 1998), p. 38.
[21] K. P. Köstli, M. Frenz, H. Bebie, and H. P. Weber, "Temporal backward projection of optoacoustic pressure transients using Fourier transform methods," Phys. Med. Biol. **46**(7), 1863–1872 (2001).
[22] D. R. Bacon, "Characteristics of a PVDF membrane hydrophone for use in the range 1–100 MHz," IEEE Trans. Sonics Ultrason. **SU-29**, 18–25 (1982).
[23] P. C. Beard, F. Perennes, and T. N. Mills, "Transduction mechanisms of the Fabry–Perot polymer film sensing concept for wideband ultrasound detection," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**(6), 1575–1582 (1999).
[24] M. J. S. Lowe, "Matrix techniques for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(4), 525–542 (1995).
[25] B. W. Suter and R. A. Hedges, "Understanding fast Hankel transforms," J. Opt. Soc. Am. A **18**, 717–720 (2001).
[26] M. J. Cree and P. J. Bones, "Algorithms to numerically evaluate the Hankel transform," Comput. Math. Appl. **26**, 1–12 (1993).
[27] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. (Cambridge University Press, Cambridge, 1999).
[28] L. C. Andrews and B. K. Shivamoggi, *Integral Transforms for Engineers and Applied Mathematicians* (Macmillan, New York, 1987).
[29] B. T. Cox, J. Laufer, K. Köstli, and P. Beard, "Experimental validation of photoacoustic *k*-space propagation models," Proc. SPIE **5320**, 238–248 (2004).
[30] G. Paltauf, H. Schmidt-Kloiber, and M. Frenz, "Photoacoustic waves excited in liquids by fiber-transmitted laser pulses," J. Acoust. Soc. Am. **104**(2), 890–897 (1998).

An aeroacoustically driven thermoacoustic heat pump[a)]

W. V. Slaton
Department of Physics and Astronomy, The University of Central Arkansas, 201 Donaghey Avenue, Conway, Arkansas 72035-0001

J. C. H. Zeegers
Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 15 September 2004; revised 23 February 2005; accepted 18 March 2005)

The mean flow of gas in a pipe past a cavity can excite the resonant acoustic modes of the cavity—much like blowing across the top of a bottle. The periodic shedding of vortices from the leading edge of the mouth of the cavity feeds energy into the acoustic modes which, in turn, affect the shedding of the next vortex. This so-called aeroacoustic whistle can excite very high amplitude acoustic standing waves within a cavity defined by coaxial side branches closed at their ends. The amplitude of these standing waves can easily be 20% of the ambient pressure at optimal gas flow rates and ambient pressures within the main pipe. A standing wave thermoacoustic heat pump is a device which utilizes the in-phase pressure and displacement oscillations to pump heat across a porous medium thereby establishing, or maintaining, a temperature gradient. Experimental results of a combined system of aeroacoustic sound source and a simple thermoacoustic stack will be presented. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1904423]

PACS numbers: 43.35.Ud, 43.28.Ra, 43.20.Ks, 43.25.Vt [RR] Pages: 3628–3635

I. INTRODUCTION

The key phenomena that has been investigated in this paper includes the coupling of aeroacoustics and thermoacoustics. While these areas of research are thoroughly studied and understood in separate experimental or practical environments, it was previously unknown how these concepts would work together. This section continues with the introduction of the fundamental concepts of aeroacoustic sound generation and thermoacoustic heat pumping. Section II details the experimental setup used in this investigation. Results of the investigation are found in Sec. III with discussion of free and forced-convection in Sec. IV. Conclusions are found in Sec. V.

A simple example of aeroacoustics is the sound generated when blowing across the top of an empty soda bottle. Sound excitation in closed side-branches due to mean flow over the side-branch openings has been extensively studied.[1–4] The shear viscous boundary layer detaches at the mouth of the side-branch when the airflow goes over the side-branch opening. This detached shear layer coils up in the form of a vortex as it is carried across the opening. This vortex feeds energy into the acoustic modes of the side-branch, which affect the shedding of the next vortex. In this way a feedback loop is established and the system is highly resonant.[5] Methods of eliminating, dampening, and predicting these undesirable acoustic waves in industrial environments have been an ongoing research goal for the aeroacoustic community. High amplitude pressure waves in gas networks can lead to fatigue in pipes, valves[6] or welds. These high amplitude pressure waves also result in spurious mean flow measurements or equipment shut-down due to excessive vibration. However, in a properly designed resonator these high amplitude pressure oscillations may be used to serve a useful purpose. Acoustic waves that propagate parallel to a solid surface are known to pump heat along the wall. If the surface is made of a low-thermal conductivity material then a temperature gradient can be established along the surface's length. This phenomena is known as thermoacoustics.

There are two types of thermoacoustic engines. The first, called a prime mover, uses thermal energy to produce useful work in the form of sound. The second, called a heat pump or refrigerator, uses work in the form of sound to pump heat over a temperature gradient. The thermoacoustic engines just described operate by transferring heat between the working media (typically a gas) and a porous structure known as a ''stack'' in the thermoacoustic literature. With appropriate thermal contact between gas and stack a significant temperature difference may be generated across the stack in the case of a refrigerator. Thermoacoustic refrigerators have been designed and built to reach cryogenic temperatures for liquefying natural gas using a traveling acoustic wave or Stirling engine cycle,[7,8] for use as a refrigerator aboard the U.S. space shuttle[9] and to cool radar electronics aboard a U.S. Navy vessel[10,11] as well as ice cream.[12] Experimental study of efficient acoustic heat pumps and their optimization continues to this day.[13]

II. EXPERIMENTAL APPARATUS

This section details several experimental setups that have been used to investigate physical acoustic phenomena in the areas of aeroacoustics and thermoacoustics. A high-pressure flow setup at the Technical University at Eindhoven was utilized to study the aeroacoustic behavior of closed-end

[a)]Portions of this work have been presented in: W. V. Slaton and J. C. H Zeegers, ''An aeroacoustically driven thermoacoustic heat pump,'' J. Acoust. Soc. Am. 113(4) Pt. 2 2269 (2003).

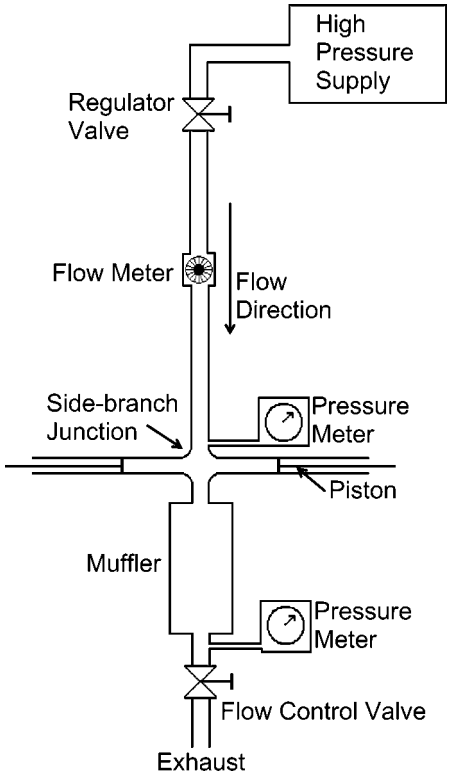


FIG. 1. Experimental flow setup used for these tests. The high-pressure supply can provide ~100 bar of air. The flow velocity of the air through the setup is controlled by the valve near the exhaust to room pressure at the bottom of the figure. The side-branch length may be adjusted with sliding pistons.

coaxial cylindrical side-branches. This phenomena is examined with empty side-branches as well as when a thermoacoustic stack is present. The high-pressure flow setup is first described including details of the aeroacoustic whistle and side-branches. Following these descriptions is an overview of the types of thermoacoustic stacks used in this investigation.

A. High-pressure flow setup

Of interest in this investigation is the configuration illustrated in Fig. 1. As indicated in the figure, the flow system is instrumented with a flow meter, several mean pressure meters as well as interchangeable flow control valves with varying flow resistances. A temperature probe at the side-branch junction measures the temperature of the flowing gas. The gas temperature varies between 22 and 18 °C for most short experimental runs but can dip down to 13 °C during extended use. This drop in temperature is due to the expansion of the gas from room temperature at high pressure (100 bars) to 10 bars in the flow setup. The muffler acts to dampen escaping higher harmonic acoustic waves generated by the whistle as well as to isolate the whistle from sound generated by flow through the flow control valve. The side-branch junction is the aeroacoustic whistle. The closed side-branches act as a resonating chamber driven by vortex shedding. The length of the side-branches may be varied independently by moving the indicated pistons. All side-branch lengths will be equal for a given half-wavelength resonance mode in the data presented below.

A pictorial representation of the half-wavelength acoustic mode in the side-branch system is illustrated in Fig. 2. The acoustic pressure is 90° out of phase with the acoustic velocity, i.e. standing wave phasing. The acoustic pressure within the co-axial side-branch resonators is given by,

$$p(z) = p(L_b)\sin[k_o z], \tag{1}$$

where k_o is the wave number of the side-branch, $k_o = 2\pi/\lambda = \pi/2L_b$, the side-branch length measured from the center of the main pipe is L_b, and z is measured from the center of the main pipe to the location of interest within the side-branch. L_b ranges from 0.57 m for 150 Hz to 0.19 m at 450 Hz. For this work the stack was positioned between 4 mm and 15 mm from the end of the side-branch, however, only the 10 mm spacing will be reported. The maximum acoustic pressure, $p(L_b)$, excited by aeroacoustic phenomena is measured by a model 116A PCB pressure transducer at an antinode of the excited mode as illustrated in Fig. 2. The acoustic velocity at the junction is inferred from, $u(0) = p(L_b)/\rho c$, where $p(L_b)$ is the aforementioned maximum acoustic pressure amplitude, ρ is the ambient gas density, and c is the ambient speed of sound. Thus, the acoustic velocity may be defined with standing wave phasing as,

$$u(z) = u(0)\cos[k_o z] = \frac{p(L_b)}{\rho c}\cos[k_o z]. \tag{2}$$

Equations (1) and (2) anticipate that the acoustic pressure and velocity profiles are not influenced by the presence of a thermoacoustic stack in the side-branch. The dimensionless ratio of acoustic velocity to main flow velocity, $u(0)/U_o$, at the junction is a useful and common method of displaying aeroacoustic sound amplitude. Bruggeman *et al.*[1,14] distinguish these regimes:

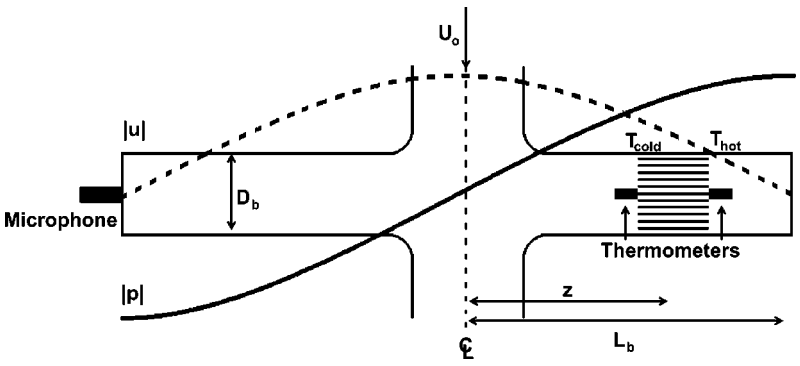


FIG. 2. Schematic detail of the side-branch system. The mean flow of air, U_o, passes past the open mouths of the coaxial side-branches. The acoustic pressure, $|P|$, (solid line) and velocity, $|u|$, (dashed line) for the half-wavelength acoustic mode within the side branches is illustrated during one instant of the acoustic cycle. The parallel lines in the right side-branch denote a thermoacoustic stack which is instrumented with thermometers.

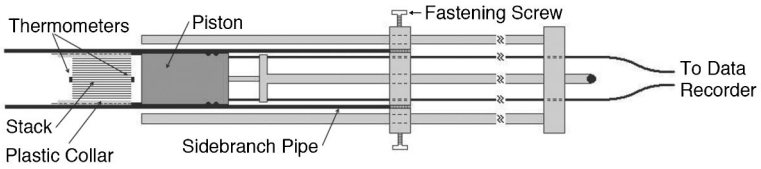


FIG. 3. Details of the Stack-Piston setup. The stack is held tightly within the side-branch pipe by a plastic collar. The piston position within the side-branch may be changed thereby altering the resonance frequency of the side-branch. The piston is held in place by fastening screws.

(1) $u(0)/U_o < 10^{-3}$: small acoustic amplitude, the growth of disturbances in the shear layer can be described by theory that is linear in the acoustic amplitude.
(2) $10^{-3} < u(0)/U_o < 10^{-1}$: moderate amplitude, the linear theory is valid for the initial shear layer disturbances. Nonlinear effects begin that limit the growth of shear layer perturbations.
(3) $u(0)/U_o = O(1)$: high amplitude, vortex generation and power of sound source strongly influenced by amplitude of acoustic field.

The dimensionless acoustic velocity amplitude will be $O(1)$ in this investigation. Due to this high amplitude the acoustic source power driving the resonance is not well characterized by theory. Modeling the performance of the aeroacoustic whistle driving a thermoacoustic stack is difficult without a good theoretical model of the available source power as a function of amplitude. Thus, for this initial investigation we will present a qualitative description of the system's performance.

The Strouhal number is a dimensionless number that is important for characterizing aeroacoustic phenomena. The Strouhal number can be thought as a dimensionless ratio of the time for the main pipe flow to pass the side-branch to the period of the acoustic disturbance. Thus,

$$S_t = \frac{f W_{\text{eff}}}{U_o}, \tag{3}$$

where f is the frequency of the periodic phenomena (the shedding of vortices or equivalently the frequency of the acoustic wave), W_{eff} is the effective diameter of the side-branch opening to the main pipe, and U_o is the mean flow velocity of gas in the main pipe grazing over the side branch opening. The effective diameter of the circular side-branch pipe with rounded edges joining the circular main pipe is defined as,

$$W_{\text{eff}} = \frac{\pi}{4} D_b + R_{\text{edge}}, \tag{4}$$

with D_b as the diameter of the side-branch and R_{edge} is the radius of curvature of the upstream edge of the junction.[2] The mouth of the side-branches are rounded such that $R_{\text{edge}}/D_b = 0.2$. The main pipe has a diameter of 30 mm and the polished stainless steel side-branches have an inner diameter of 25 mm and a wall thickness of 5 mm. Within this pipe a brass piston with o-rings can slide to set the side-branch length, see Fig. 3. The piston is held in place by steel gliders on the outside, which are held by fastening screws to the steel pipe.

The interest in this investigation is to use this high-amplitude sound source as a driver for a simple thermoacoustic heat pump. The acoustic power available from the vortex shedding process serves to overcome viscous/thermal losses within the side-branch resonators as well as high-amplitude nonlinear wave steepening effects which lead to higher harmonic generation. When a thermoacoustic stack is introduced in the side-branch the power in the acoustic wave overcomes viscous and thermal losses within the stack as well as pumping heat from one end of the stack to the other. Due to the lack of heat exchangers and the lack of knowledge of the high-amplitude source power of the vortex-shedding phenomena as a function of mean flow, frequency and acoustic amplitude, the experimental results can at best be presented as a qualitative rather than quantitative, description. Nevertheless, the qualitative description will prove useful to future work utilizing an aeroacoustic sound source in conjunction with thermoacoustic effects.

B. Thermoacoustic stacks

A conceptual sketch of how the aeroacoustic whistle is used with a thermoacoustic stack is also illustrated in Figs. 2 and 3. Experiments were always done with one stack in the side-branch opposite the microphone as displayed in the figures. To record the thermoacoustically generated temperatures, PT-1000 platinum temperature resistors are fastened to the center of each end of the stacks using a high temperature silicone adhesive. Thin steel tubes go through the brass plunger and are sealed with tiny O-rings to allow electrical wires to connect the stack thermometers to external data recorders. Gently pushing or pulling on these thin tubes allows the stack position within the side-branch to be changed even when the system is under pressure. There are three critical thermoacoustic dimensions, the length of the stack, the position of the stack in the side-branch and the characteristic open pore dimension in the stack. To investigate these parameters, two types of stacks have been constructed. The first type consists of stacked parallel plates made of Mylar plastic with varying plate spacing, however, the results of these Mylar stacks will not be presented due to space limitations. The second type of stack consists of square pore ceramic material (obtained from Dow Corning and NGK Europe GmbH) of various pore size cut to varying lengths. Both types of stack are held firmly in a 1.5 mm thick, 10 cm long plastic collar with paper padding for a snug fit in the side-branch pipe. The details of the dimensions of the ceramic stacks are summarized in Table I with the thermophysical properties of the ceramic given by its thermal conductivity, ~ 2 W/m K, its bulk density, 217 kg/m^3, and its specific heat capacity, 889 J/kg K, also note that one mil is equivalent to 2.54×10^{-2} mm.

A thermoacoustic engine with a temperature gradient across its stack greater than a critical value known as a prime-mover while a heat-pump has a temperature gradient

TABLE I. Physical properties of the stacks under investigation.

Ceramic stack material (CPSI)	Stack wall thickness (mil)	Stack length (mm)	Stack diameter (mm)	Hydraulic diameter of pore (mm)	Total porosity (unitless)
900	2.5	20.12	21.78	0.783	67.9%
900	2.5	40.01	21.70	0.783	68.0%
900	2.5	60.64	22.43	0.783	67.2%
900	2.5	80.45	21.69	0.783	68.0%
900	2.5	100.59	22.08	0.783	67.6%
300	8.0	39.10	21.90	1.263	59.1%
400	4.0	40.00	21.80	1.168	67.2%
600	3.0	39.65	22.05	0.961	67.8%
900	2.5	39.95	21.70	0.783	67.8%
1200	2.5	39.55	21.85	0.670	66.2%
1500	2.5	39.80	21.60	0.592	65.1%
1500	2.5	60.05	21.80	0.592	64.8%

less than this critical value. The inviscid, standing wave expression for the critical gradient is given by,[15]

$$\nabla T_{\mathrm{crit}} = \frac{T_o \beta \omega}{\rho c_p} \frac{p(z)}{u(z)}, \tag{5}$$

where T_o is the ambient temperature, β is the volumetric expansion coefficient of the gas ($\beta = 1/T_o$ for an ideal gas), ω is the angular frequency, ρ is the ambient density, c_p is the isobaric specific heat, and z is measured from the center of the main pipe to the center of the stack with $p(z)$ and $u(z)$ defined previously. The equation for the critical gradient shows that the mean temperature within the stack should be increasing in the $+z$-direction, i.e. from the side-branch junction to the closed end. Thus, the cold side of the stack faces the junction with the hot side facing the closed end of the side-branch. The critical gradient is largest for stacks placed near an acoustic velocity node. This means that larger temperature gradients will be generated across stacks placed near the closed end of the side-branch. Also, the ratio of the measured temperature gradient to the critical gradient gives a useful dimensionless ratio, Γ.

A critical dimension within the stack is the size of the pores. Optimal performance for standing wave systems results when the hydraulic radius (equal to the pore cross-sectional area divided by the pore perimeter) of the pore is 0.5–1.25 thermal penetration depths for parallel plate, rectangular and circular pore geometries.[15–18] The hydraulic radius (R_h) for square pores of dimension ($a \times a$) is $a/4$. The thermal penetration depth can be thought of as the distance heat can diffuse through the gas in a time $1/\omega$. The thermal penetration depth is defined as, $\delta_\kappa = \sqrt{2\kappa/\omega\rho c_p}$, where κ is the gas thermal conductivity and the other symbols are the same as defined before. Viscous effects, imposed by the no-slip boundary condition at solid walls, also diffuse through the gas with a penetration depth, $\delta_v = \sqrt{2\mu/\omega\rho}$ with μ as the gas viscosity.

For high amplitude acoustic waves in stacks it is convenient to investigate phenomena as a function of an acoustic Reynolds number similar to one defined by Swift[18] given by,

$$|\mathrm{Re}_{\mathrm{ac}}|_{\max} = \frac{|u_{\mathrm{ac}}|_{\max} D_h \rho}{\mu}, \tag{6}$$

where D_h is the hydraulic diameter (equal to four times the hydraulic radius) of the pore, ρ is the ambient density, μ is the ambient viscosity, and $|u_{\mathrm{ac}}|_{\max}$ is the maximum acoustic velocity at the location of interest in the resonator. For this investigation, the location of interest is the stack center. The acoustic Reynolds number may also be defined with respect to the viscous penetration depth as opposed to the pore hydraulic diameter.[19] Another important parameter, which includes the frequency of the acoustic wave, is the ratio of the hydraulic diameter to the viscous penetration depth, D_h/δ_ν. For circular pipes, Swift[18] notes that a laminar to turbulent transition occurs roughly at an acoustic Reynolds number of 2000 for $D_h/\delta_\nu < 10$. Above $|\mathrm{Re}_{\mathrm{ac}}|_{\max} = 2000$ the fluid motion can be weakly turbulent (turbulence is confined to the center of the pore), conditionally turbulent (flow transitions from laminar to weakly turbulent during acoustic cycle) or completely turbulent. An interesting plot of the regions of turbulence as a function of D_h/δ_ν and $|\mathrm{Re}_{\mathrm{ac}}|_{\max}$ is illustrated by Swift in his text. We will use this framework when analyzing our experimental results.

All stacks undergo the same experimental procedure. After the stack is installed within the side-branch and secured, the mean pressure within the flow setup may be adjusted. As the flow control valve is opened the mean flow of gas past the side-branches excites the half-wavelength acoustic mode over a range of Strouhal numbers. The aeroacoustically generated sound interacts with the stack to develop a temperature difference. The mean pressure and temperature of the flowing gas as well as its flow rate is recorded along with the maximum acoustic pressure amplitude, $p(L_b)$, in addition to the stack temperatures. For each stack the flow rate is swept over the region of interest at each frequency (side-branch length) of interest, not all of this data will be presented however.

A word of caution should be noted at this point. The stacks are not fitted with heat exchangers, however, the thick walled steel pipe acts as a large thermal reservoir as well as a heat bridge across the stack. The stack/side-branch system takes a long time to equilibrate (hours) and the flow setup of the Technical University at Eindhoven cannot provide sustained high-velocity airflow under high-pressure for more than a few hours per day. Thus, to speed up data collection the experiment is done fast enough to allow the faces of the stack to equilibrate but not the side-branch pipe. All data for a particular variable setting takes approximately 1 min to record and this is enough time for the stack to achieve equilibrium with itself. The stack and side-branch are allowed to cool down completely after each sweep of the flow rate at a given frequency; which takes approximately 10 min.

III. EXPERIMENTAL RESULTS

This section contains experimental results obtained with the experimental apparatus described in the previous section. First an examination of the aeroacoustic sound generation in the empty steel side-branches is given. This is followed by results obtained with the square-pore ceramic stacks. In gen-

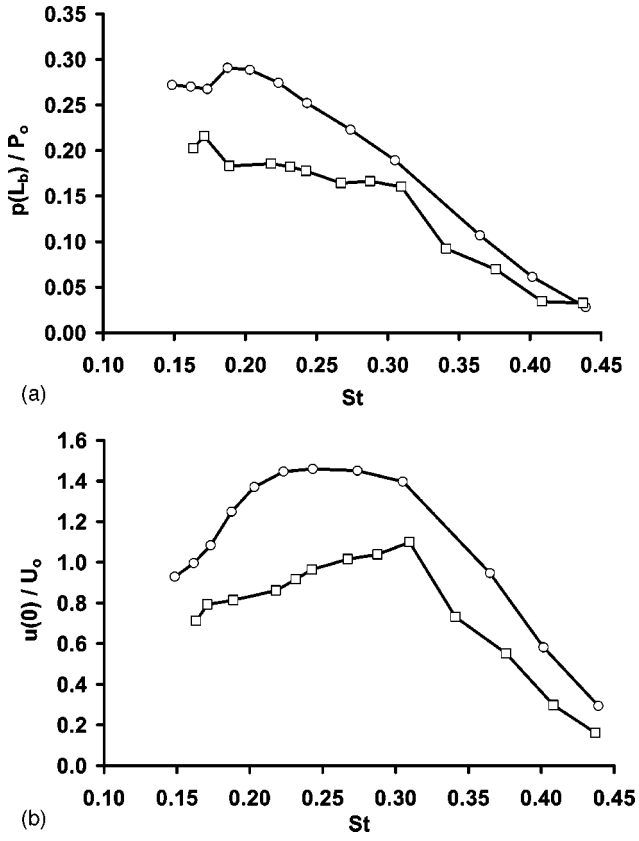


FIG. 4. (a) Dimensionless acoustic pressure amplitude versus Strouhal number for a frequency of ~350 Hz with no stack present at 2 and 10 bar ambient pressure. Circles are for $P_o = 10$ bar and squares are for $P_o = 2$ bar. (b) Dimensionless acoustic velocity amplitude versus Strouhal number for a frequency of ~350 Hz with no stack present at 2 and 10 bar ambient pressure. Circles are for $P_o = 10$ bar and squares are for $P_o = 2$ bar.

eral, the absolute pressure is 10 bars unless otherwise noted and the half-wavelength acoustic mode is excited in the side-branches. Also, the lines between data points in plots are merely guides for the eye.

The experimental results to be presented in this section are a subset of the total data collected. Results at a fixed frequency (side-branch length) will be displayed as a function of the Strouhal number. This variable flow velocity data helps to understand the aeroacoustic whistle's performance both with and without a stack present. At other frequencies the same experimental procedure of sweeping the mean flow rate has been repeated, however, these data are not presented.

A. Aeroacoustic whistle performance

The performance of the aeroacoustic whistle for the half-wavelength acoustic mode for the empty side-branches at mean pressures of 2-bar and 10-bar and a fixed frequency of ~350 Hz is displayed in Figs. 4(a) and (b). The measured acoustic pressure amplitude relative to the mean pressure is plotted as a function of the Strouhal number in Fig. 4(a), while the maximum acoustic velocity in the side branch junction relative to the mean flow velocity is plotted as a function of the Strouhal number in Fig. 4(b) using, $u(0) = p(L_b)/\rho c$. Since the frequency is essentially constant (varying from 348 Hz at low mean flow velocity to 356 Hz at high velocity) and the size of the mouth of the side-branch does not change, the Strouhal number in these figures represents the changing mean flow velocity. The plot of the acoustic velocity to mean flow velocity shows a peak near a Strouhal number of 0.25 at 10 bar corresponding to a flow velocity of ~35 m/s. The side-branches resonate in the half-wavelength mode only over the indicated range of Strouhal numbers. Near a Strouhal number of 0.15 the resonance is unstable and can collapse abruptly to zero amplitude. Note that the maximum dimensionless acoustic pressure amplitude is 30% of the mean pressure at 10 bars, this translates to a 3 bar real pressure oscillation.

B. Square pore ceramic stacks

Ceramic stacks are cut to 2, 4, 6, 8, and 10 cm lengths from a sample with 900 square cells per square inch (CPSI). Next, stacks of 4 cm length are cut from ceramic samples with pore sizes ranging from 300 to 1500 CPSI. These stacks have varying pore size that allow for observing the effect on the thermoacoustic heat pumps' performance. Lastly, a 6 cm long stack with 1500 CPSI is tested at high acoustic amplitude for several hours leading to very large temperature differences generated across the stack.

1. 900 CPSI stacks with various length

The 900 CPSI stacks of length 2, 4, 6, 8, and 10 cm are first tested at a fixed frequency of 300 Hz and a variable flow rate, see Figs. 5(a) and (b). The stacks are placed at 5 mm, 10 mm, and 15 mm from the end of the side-branch, however, only the 10 mm spacing will be presented. The dimensionless acoustic pressure amplitude versus Strouhal number appears in Fig. 5(a). These data show how important stack length is to the overall damping of the acoustic resonance. The detection of an acoustic resonance in the side-branches occurs at a larger Strouhal number for the 2, 4, and 6 cm stacks compared to the no-stack data. The reason for this could be due to the additional acoustic damping caused by the presence of the stack. This damping affects all acoustic fluctuations in the pipe but may allow the fluctuation corresponding to the half-wavelength resonance to be detected at a lower mean flow velocity. The 10 cm stack completely damped the resonance when placed 10 mm from the closed end of the side-branch and thus this stack does not appear in the results. As the Strouhal number for the mean flow past the side-branch junction decreases, the measured acoustic pressure amplitude increases for each stack length. Longer stacks have more surface area for viscous-thermal losses and thermoacoustic heat pumping, and thus result in lower amplitude compared to the no-stack data. The corresponding temperature difference across the stacks of various lengths is displayed in Fig. 5(b). As the stack length increases the measured temperature difference over the stack increases even as the acoustic pressure drops until the best performance is found for the 6 cm stack which develops a ~90 °C temperature difference.

Dimensionless numbers based upon the maximum acoustic pressure amplitude, frequency, and temperature difference may be calculated for the data in Figs. 5(a) and (b). Table II lists the maximum dimensionless acoustic pressure

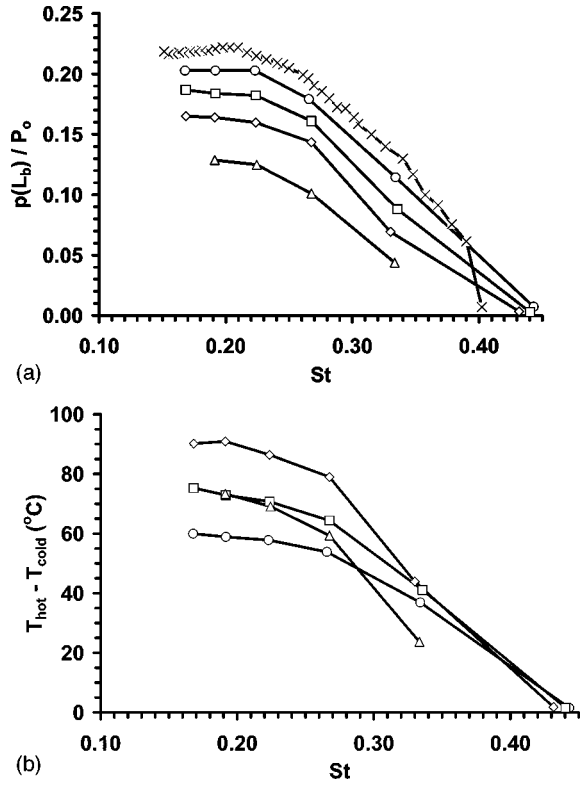


FIG. 5. (a) Dimensionless acoustic pressure amplitude at a frequency of 300 Hz with the 900 CPSI stacks located 10 mm from the end of the side-branch. Circles, 2 cm stack length; squares, 4 cm; diamonds, 6 cm; triangles, 8 cm; and × crosses denote the no-stack case. (b) Measured temperature difference across 900 CPSI stacks of various lengths at a frequency of 300 Hz with the stacks located 10 mm from the end of the side-branch. Circles, 2 cm stack length; squares, 4 cm; diamonds, 6 cm; and triangles, 8 cm.

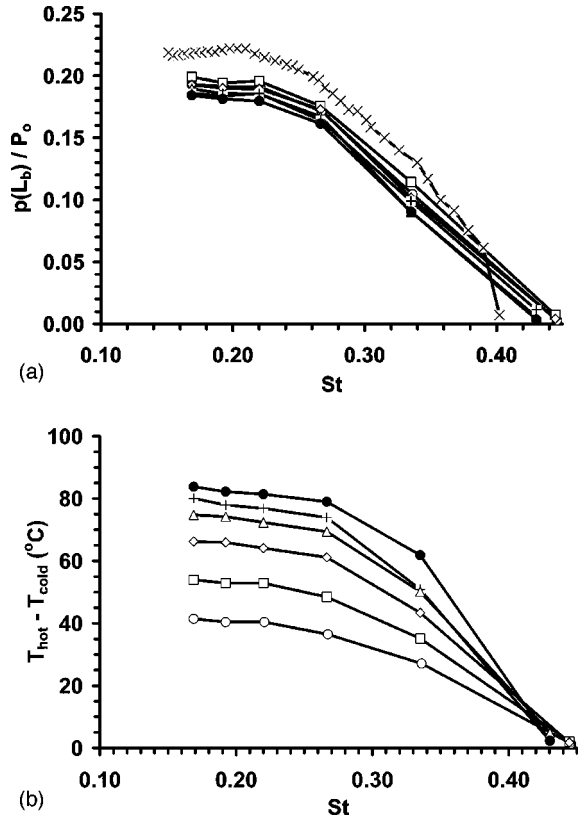


FIG. 6. (a) Dimensionless acoustic pressure amplitude at a frequency of 300 Hz with the 4 cm stacks located 10 mm from the end of the side-branch. Open circle, 300 CPSI stack; square, 400 CPSI; diamond, 600 CPSI; triangle, 900 cpsi; pluses, 1200 CPSI; closed circles, 1500 CPSI; and crosses denote the no-stack case. (b) Measure temperature difference across the 4 cm stacks of various CPSI at a frequency of 300 Hz and 10 mm from the end of the side-branch. Open circle, 300 CPSI stack; square, 400 CPSI; diamond, 600 CPSI; triangle, 900 cpsi; pluses, 1200 CPSI; closed circles, 1500 CPSI.

amplitude, $p(L_b)/P_o$, the corresponding dimensionless acoustic velocity amplitude, $u(0)/U_o$, at the junction as well as the maximum acoustic Reynolds number at the stack center, the hydraulic diameter to the viscous penetration depth, D_h/δ_ν, as well as the hydraulic radius to the thermal penetration depth, R_h/δ_κ, and the ratio of the temperature gradient to the critical gradient, Γ. If the approximate boundaries between regions of flow instabilities for oscillating flow in smooth circular pipes are roughly equivalent for the square pores of these stacks, then the oscillating gas at the center of the stacks is probably in the weakly turbulent regime. Also, the ratio of hydraulic radius to thermal penetration depth is larger than optimal and yet significantly large temperature differences are measured across the stacks.

TABLE II. Dimensionless results for stacks driven at high-amplitude.

Ceramic stack material (CPSI)	Stack length (mm)	$u(0)/U_o$	$p(L_b)/P_o$	$\lvert\mathrm{Re}_{ac}\rvert_{max}$	D_h/δ_ν	R_h/δ_κ	Γ
900	20.12	1.08	0.20	3536	18.0	3.8	0.66
900	40.01	0.99	0.19	4613	17.6	3.7	0.61
900	60.64	0.88	0.17	5457	17.5	3.7	0.65
900	80.45	0.78	0.13	5446	17.8	3.8	0.51
300	39.10	1.03	0.19	8291	29.6	6.2	0.36
400	40.00	1.07	0.20	7753	26.9	5.7	0.45
600	39.65	1.03	0.19	6085	22.0	4.6	0.55
900	39.95	0.99	0.19	4511	17.5	3.7	0.60
1200	39.55	1.02	0.19	4265	15.3	3.2	0.68
1500	39.80	0.99	0.18	3603	13.3	2.8	0.71
1500	60.05	1.16	0.164	2615	11.3	2.4	0.73

2. CPSI stacks with 4 cm length

The 4 cm stacks of 300, 400, 600, 900, 1200, and 1500 CPSI are tested at a fixed frequency of 300 Hz and a variable flow rate, see Figs. 6(a) and (b). The pore hydraulic diameter ranges from 1.2 mm at 300 CPSI down to 0.6 mm at 1500 CPSI. The stacks are placed 10 mm from the end of the side-branch. The dimensionless acoustic pressure amplitude versus Strouhal number appears in Fig. 6(a). This data shows how important pore size is to the overall damping of the acoustic wave. Smaller pore stacks draw more power out of the wave to overcome losses and thus result in lower amplitude compared to the no-stack case. However, the damping due to viscous losses is a small effect because the stacks are all 4 cm in length and are positioned near the acoustic velocity node at the end of the side-branch. Again, the aeroacoustically excited resonance begins at a larger Strouhal number than the no-stack case because of the additional acoustic damping caused by the presence of the stack. The

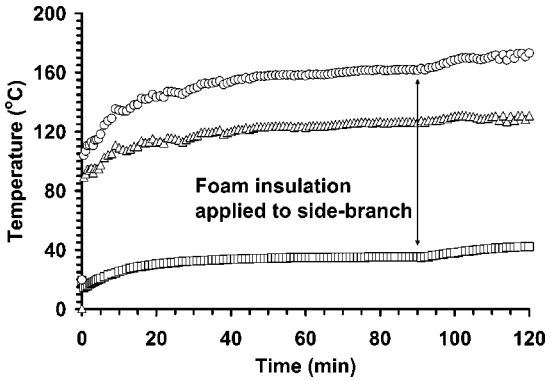


FIG. 7. Time elapse temperature measurements of a 6 cm long 1500 CPSI stack located 4 mm from the end of the side-branch. Circles denote hot side temperature; squares, cold side; and triangles, the temperature difference. Foam insulation was wrapped around the exterior of the steel side-branch pipe at the 90 min mark.

corresponding temperature difference across these stacks is displayed in Fig. 6(b). The data show that increased temperature differences across the stacks occurs for larger amplitudes as well as smaller pore stacks. The saturation of the temperature difference at high amplitude is due to thermal conduction losses through the steel side-branch pipe and convection from the exterior of the side-branch into the room.

Table II lists dimensionless values for the data in Figs. 6(a) and (b). Again, assuming that the boundaries for laminar and turbulent oscillating flow for circular pipes are comparable to our square-pore stacks it is seen that the oscillating flow in the stacks in this data set are also in the weakly turbulent regime at their maximum acoustic amplitudes. The largest temperature difference is for the 1500 CPSI stacks which have $R_h/\delta_\kappa \sim 2.8$ at this frequency which is still not optimal. However, even with this limitation, the 1500 CPSI stack attains a temperature difference of approximately 85 °C at the largest amplitudes.

3. Time elapse

Based upon the results of the 900 CPSI stacks of various length and the 4 cm stacks of various CPSI, a stack of 6 cm length was cut from the 1500 CPSI ceramic. This stack was tested at a fixed flow rate of 34 m/s at 317 Hz with a Strouhal number of 0.22. The acoustic amplitude was 16% of the mean pressure of 10 bars which resulted in a maximum acoustic Reynolds number of ~ 2615 at the stack center. The stack was placed 4 mm from the closed end of the side branch and had a ratio of hydraulic radius to thermal penetration depth of ~ 2.6 at this operating point. Data was recorded for this configuration for 2 h and can be seen in Fig. 7. The figure displays the hot and cold side temperatures as well as the temperature difference as a function of the time. During the first 30 s a rapid increase in temperature at the hot side occurs, after that the hot side temperature steadily rises until reaching a fairly steady value of 160 °C ($\Gamma = 0.73$) after 90 min. The cold side of the stack begins at room temperature and heats up to 35 °C due to conduction and convection of heat from the hot side. After 90 min foam insulation was placed around the steel side-branch pipe to insulate it from the room, both hot and cold side temperatures rise because the heat that was leaving the side-branch due to free convection in the room had been eliminated. The temperature difference across this stack is also plotted in Fig. 7. It can be seen that the stack begins with 90 °C temperature difference very quickly after the sound is excited, this temperature difference continues to grow even after the foam insulation is applied. This is a remarkably large temperature difference for such a simple thermoacoustic heat pump. This test shows that large temperature differences in excess of 100 °C can be generated across a thermoacoustic stack, which is placed near the closed end of a aeroacoustically excited resonator. The authors know of no other experiment with a standing-wave thermoacoustic system that has attained such large temperature differences.

IV. CONDUCTION AND CONVECTION

Experiments have also been carried out with this setup to investigate thermal conduction in the side-branch as well as free and forced convection. To investigate thermal conduction in the side-branch pipe as a possible source for losses, an identical side-branch was constructed of plexiglass with 2.5 mm wall thickness. The plexiglass has a low thermal conductivity, 0.2 W/mK compared to stainless steel's 14 W/mK. However, because the stacks in this investigation were not fitted with heat exchangers the viscous and thermal energy dissipated in the stack simply increased the mean temperature of the stack and did not result in larger temperature differences.

Free convection resulting from temperature driven buoyancy effects could result in hot gas flowing through the stack and thereby leading to losses. To investigate the presence of buoyancy driven free convection the steel side-branches were rotated with respect to the main pipe. However, no measurable differences in performance of the stacks were observed with the hot side of the stack rotated to an angle of $+39°$ above the horizontal or to $-58°$ below the horizontal. These particular angles are the result of the range of rotation allowed by the main pipe's height above the floor and the length of the side-branch arms. Buoyancy driven free convection was also investigated by affixing eccentrically placed thermometers on the hot side of the stack at the stack center, halfway to the edge, and at the edge of the stack. Measurements with these eccentrically placed thermometers in the steel pipe also failed to show free convection effects but did show a temperature gradient from the stack center to the side-branch wall which is not surprising given that the stacks were not completely in thermal equilibrium with the side-branch.

Lastly, the turbulent flow of gas in the main pipe probably excites large forced convection flows in the side-branches. These forced convection flows were not visualized in the experimental setup but were observed in a similar setup using in-house compressed air and smoke. Placing a thin wire gauze within the side-branch between the stack and the junction with the main pipe to break up the forced convection flows only served to drastically dampen the excited acoustic wave and did not result in enhanced temperature differences across the stack.

V. CONCLUSIONS

Experimental tests have been carried out which couple a simple thermoacoustic stack to a high-amplitude aeroacoustic sound source. An aeroacoustic sound source has never been coupled to a thermoacoustic device to pump heat and establish a temperature gradient. This method of resonance excitation provides challenges not encountered when using electrodynamic drivers, however, the aeroacoustic source will support an acoustic load and operates with no moving parts. These experimental tests have demonstrated that the length of the stack and the ratio R_h/δ_κ is important to a thermoacoustic heat pump's performance even when operating in or near a turbulent regime. The thermoacoustically generated temperature differences are quite large for such a simple device and could be used in conjunction with thermoelectric elements to generate electric power. Specifically, such a power generation scheme could be employed in the high-pressure, high-velocity, natural gas environment found down-well or in pipelines. Proper design of a no-moving-parts electrical power generator for down-well applications must account for: the heat load caused by the presence of the thermoelectric element and improving the geometry of the setup to fit within a well pipe as well as potential clogging of the stack pores by debris. Optimizing energy transfer from the initial hydrodynamic flow to the thermoelectric element is another avenue for improvement. It should also be noted that unwanted aeroacoustic excitations in duct networks with flow may be dampened or eliminated by the inclusion of a porous material near a region of high acoustic velocity amplitude for the excited mode. Further work to accurately characterize the available source power of the aeroacoustic sound source at high acoustic amplitudes is needed to fully characterize the system.

ACKNOWLEDGMENTS

The authors would like to thank the following students who participated in this project: Pier Dolmans, Rob Tummers, and Hans Crijns. The authors would also like to thank Jan Willems for constructing the high-pressure flow setup and Avraham Hirschberg for aeroacoustic expertise, which was invaluable for this project to succeed. Lastly, the authors appreciate the support of Alex van der Spek who conceived of this idea for electrical power generation and who supplied funding through Shell's International Exploration and Production B.V. department. The authors would also like to thank the reviewers for helpful suggestions and comments.

[1] Jan C. Bruggeman, "Flow induced pulsations in pipe systems," Ph.D. dissertation, Applied Physics Department, Technical University at Eindhoven, The Netherlands, 1987.

[2] M. C. A. M. Peters, "Aeroacoustic sources in internal flows," Ph.D. dissertation, Applied Physics Department, Technical University at Eindhoven, The Netherlands, 1993.

[3] S. Dequand, "Duct aeroacoustics: from technological applications to the flute," Ph.D. dissertation, Applied Physics Department, Technical University at Eindhoven, The Netherlands, 2001.

[4] J. C. Bruggeman, A. Hirschberg, M. E. H. van Dongen, A. P. J. Wijnands, and J. Gorter, "Self-sustained aeroacoustic pulsations in gas transport systems: experimental study of the influence of closed side branches," J. Sound Vib. **150**, 371–393 (1991).

[5] P. C. Kriesels, M. C. A. M. Peters, A. Hirschberg, A. P. J. Wijnands, A. Iafrati, G. Riccardi, R. Piva, and J. C. Bruggeman, "High amplitude vortex-induced pulsations in a gas transport system," J. Sound Vib. **184**, 343–368 (1995).

[6] R. M. Baldwin and H. R. Simmons, "Flow induced vibration in safety relief valves," J. Pressure Vessel Technol. **108**, 267–272 (1986).

[7] G. W. Swift, "Thermoacoustic natural gas liquefier," in Proceedings of the DOE Natural Gas Conference (Fed. Energy Tech. Cent., Morgantown, West Virginia, 1997).

[8] G. W. Swift, "Efficiency Issues for Large-Scale Thermoacoustic Liquefaction of gases," First International Workshop on Thermoacoustics, s'-Hertogenbosch, The Netherlands (2001).

[9] S. L Garrett, J. A. Adeff, and T. J. Hofler, "Thermoacoustic refrigerator for space applications," J. Thermophys. Heat Transfer **7**, 595–599 (1993).

[10] S. C. Ballister and D. J. McKelvey, "Shipboard electronics thermoacoustic cooler," M.S. thesis, Naval Postgraduate School, Monterey, CA (1998).

[11] R. A. Johnson, S. L. Garrett, and R. M. Keolian, "Thermoacoustic cooling for surface combatants," Nav. Eng. J. **112**, 335–345 (2000).

[12] M. E. Poese, R. W. M. Smith, and S. L. Garrett, "Regenerator-based thermoacoustic refrigerator for ice cream storage applications," J. Acoust. Soc. Am. **114**(4), 2328 (2003).

[13] M. E. H. Tijani, "Loudspeaker-driven thermoacoustic refrigeration," Ph.D. dissertation, Applied Physics Department, Technical University at Eindhoven, The Netherlands, 2001.

[14] J. C. Bruggeman, A. Hirschberg, M. E. H. van Dongen, A. P. J. Wijnands, and J. Gorter, "Flow induced pulsations in gas transport systems: analysis of the influence of closed side branches," J. Fluids Eng. **111**, 484–491 (1989).

[15] G. W. Swift, "Thermoacoustic Engines," J. Acoust. Soc. Am. **84**(4), 1145–1180 (1988).

[16] M. E. H. Tijani, J. C. H. Zeegers, and A. T. A. M. de Waele, "The optimal stack spacing for thermoacoustic refrigeration," J. Acoust. Soc. Am. **112**, 128–133 (2002).

[17] W. Arnott, H. Bass, and R. Raspet, "General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections," J. Acoust. Soc. Am. **90**, 3228–3237 (1991).

[18] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, New York, 2002), Chap. 7.

[19] P. Merkli and H. Thomann, "Transition to turbulence in oscillating pipe flow," J. Fluid Mech. **68**, 567–575 (1975).

Acoustic characteristics of an electrodynamic planar digital loudspeaker using noise shaping technology

Atsushi Hayama and Kenji Furihata[a)]
Department of Electrical and Electronic Engineering, Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano 380-8553, Japan

David K. Asano
Department of Information Engineering, Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano 380-8553, Japan

Takesaburo Yanagisawa
Department of Electrical and Electronic Engineering, Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano 380-8553, Japan

(Received 7 July 2004; revised 26 January 2005; accepted 3 February 2005)

The present study extends our previous work [Furihata *et al.*, J. Acoust. Soc. Am. **114**, 174–184 (2003)] by investigating our electrodynamic planar loudspeaker when driven by a 12 bit digital signal with noise shaping. Changing the structure of the loudspeaker can lead to improvement, but in this paper improvements that can be made using signal processing are investigated. Results show that the digital loudspeaker demonstrated good linearity over its 84 dB dynamic range from 40 Hz to 10 kHz. This shows that a 12 bit digital loudspeaker which is equivalent to a 16 bit one is possible. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1887025]

PACS numbers: 43.38.Dv, 43.38.Ja [AJZ] Pages: 3636–3644

I. INTRODUCTION

In a recent study,[1] an electrodynamic planar digital loudspeaker was presented. In order to achieve an adequate sound pressure level and bit resolution, the loudspeaker was made with a diaphragm molded in plastic, which allows the creation of streamlined sections in order to suppress divided vibrations. Two voice coils were rolled together in one section of the diaphragm, resulting in a total of 22 voice coils. Also, 11 permanent magnets were arranged under the diaphragm so that the poles of adjacent magnets were opposite to each other. The suspension between the diaphragm and the frame was made of a piece of handmade Japanese paper.

The acoustic responses, such as frequency and distortion, were affected by the selection of the voice coils. Therefore, in order to improve the responses, (a) the combination and (b) the arrangement of the voice coils were discussed. As for (a), pairs of voice coils were chosen by analyzing the results of each acoustic response when the loudspeaker was driven by an analog signal. With regards to (b), each bit was assigned to a voice coil based on the result of the best combination in (a) resulting in an asymmetric arrangement. This asymmetric arrangement was designed to obtain as flat a frequency response to an analog signal as possible. This asymmetric arrangement was compared with a symmetric one and the results showed that the flatness of the frequency responses around 1 and 4 kHz were improved and that the sound reproduction band was from 40 Hz to 10 kHz.

In order to compare the digital loudspeaker with the above-assigned asymmetric arrangement to one with a symmetric arrangement, experiments were performed in which the loudspeakers were driven by a weighted discrete voltage with a maximum amplitude of 16 $V_{p\text{-}p}$ and a resolution of 16 bits, i.e., the same number of bits used in CD (compact disc) and DAT (digital audio tape) recording. In these experiments, the output wave form, frequency response, linearity, total harmonic distortion, distribution of sound intensity level, and directional pattern of both arrangements were evaluated. The results showed that the digital loudspeaker can reproduce pure sound with a distortion factor less of than about 5% from 40 Hz to 10 kHz, exceeding this value only in a narrow frequency band near 4 kHz. This digital loudspeaker was demonstrated to have good linearity over its dynamic range of 84 dB.

From contour maps of the sound intensity level, it was found that the diaphragm movement changes from piston motion at 3.15 kHz to divided vibration mode motion at 4 kHz. In particular, the sound intensity is highest, 86.4 dB, on the lower right surface, and lowest, 80.3 dB, on the upper left surface.

In this paper, in order to improve the acoustic responses further, all voice coils are considered when selecting an arrangement and combinations. However, the number of bits has to be reduced to 12, or the dynamic range will deteriorate because pulse code modulation cannot resolve details smaller than the least-significant bit. Hence, some interesting ideas have been proposed to try to maximize the human-auditory potential. One idea is noise-shaping. Noise-shaping was first proposed by Michael Gerzon and Peter Craven in 1989,[2] and successfully embodied in Meridian's 618,[3] 518,[4] and also in Sony's Super Bit Mapping.[5] Therefore, noise shaping technology is applied to solve this problem.

[a)]Electronic mail: kennfur@gipwc.shinshu-u.ac.jp

 0001-4966/2005/117(6)/3636/9/$22.50

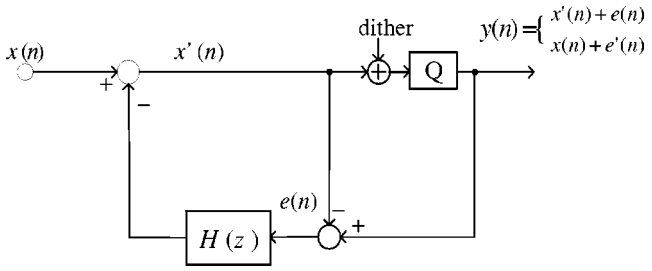


FIG. 1. Block diagram of the noise shaping filter with dithered requantization and error feedback filter $H(z)$. The requantization error is $e'(n)$.

II. NOISE SHAPING

A. Principle (Ref. 6)

Signal requantization is applied in digital audio systems whenever the word-length of audio samples needs to be reduced. This is the case for instance when an audio signal has to be stored on a CD. In entertainment audio, requantization to 8 or 12 bits could be an economically interesting alternative to other forms of data compression because requantized data can be sent directly to a D-A (digital-to-analog) converter, while encoded data needs to be decoded first. Since decoding can be computationally intensive, this could be an important advantage for using requantized data in our digital loudspeaker.

While a white noise dither signal can already improve the quality of low level requantized signals, noise shaping can additionally be applied in order to make the requantization error minimally audible.[7] Figure 1 shows the general scheme for signal requantization with noise shaping. In this scheme, Q represents the quantizer and $H(z)$ is the error feedback filter. Due to the requantization error $e(n)$, the output $y(n)$ differs from $x'(n)$ and from $x(n)$. The error feedback filter has to be controlled such that the difference between $y(n)$ and $x(n)$ becomes minimally audible.

With signals defined as shown in Fig. 1, and using z transforms, we have

$$X'(z)=X(z)-H(z)E(z), \tag{1}$$

$$X'(z)=X(z)+E'(z)-E(z), \tag{2}$$

where $E(z)$ represents quantizer Q's error signal and $E'(z)$ is the additive quantization distortion at the output of the noise shaping requantizer. Subtracting Eq. (2) from Eq. (1), we find that

$$E'(z)=\{1-H(z)\}E(z). \tag{3}$$

The requantization error $e'(n)$ therefore has a power spectrum given by

$$P_{E'}(e^{j\omega})=\|1-H(e^{j\omega})\|^2 P_E(e^{j\omega}). \tag{4}$$

As we can see, the quantizer's error spectrum gets shaped by a noise shaping filter that depends on the error-feedback filter $H(z)$.

This result was obtained without reference to the quantizer Q. Therefore, it applies to any type of quantizer, whether it is dithered, linear, nonlinear, uses rounding, or uses truncating.

In order to achieve minimal audibility of the requantization error, $H(e^{j\omega})$ can be designed to minimize the total amount of perceptually weighted noise power N_ω:

$$N_\omega=\int_{-\pi}^{+\pi} P_{E'}(e^{j\omega})W(\omega)d\omega, \tag{5}$$

where $W(\omega)$ is a perceptual weighting function that approximates the relative audibility of noise power at the different frequencies.

Super Bit Mapping (SBM)[5] introduces a clever trick to design a minimum phase FIR (finite impulse response) noise shaping filter with a given power spectral shape. Note that in order to avoid delayless loops in Fig. 1, it is required that the FIR noise shaping filter can be written as

$$1-H(z)=\sum_{n=0}^{M} a(n)z^{-n}$$

where

$$a(0)=1. \tag{6}$$

It was observed in Ref. 5 that an Mth-order inverse LPC (linear predictive coding) filter is minimum phase (guaranteed if obtained from the autocorrelation formulation[8]) and satisfies Eq. (6). Thus, the required minimum phase FIR noise shaping filter can be obtained by approximating the inverse of the desired noise shaping spectrum with a LPC synthesis filter and inverting the result.

In SBM, the desired noise shaping spectrum is taken to be the hearing threshold when no audio is present. Although SBM can be successfully applied to make the quantization error minimally audible in the absence of audio, spectral shaping that minimizes the audibility of the requantization error in the presence of the actual audio is preferable.

B. Noise shaping filter designed procedure

The noise shaping filter was designed as follows.

(1) In order to take into account all voice coils when selecting the arrangement and the combinations, the number of bits used in the dithered quantizer (Q) shown in Fig. 1 was set to 12.
(2) The noise shaper is unstable if the noise transfer function gain in the stop band is too large. Therefore, a software version of the noise shaping requantizer (Fig. 1) was implemented. A 48 kHz sampling frequency and 12th order FIR designs for $H(z)$ were used to avoid oscillation.[5]
(3) The filter coefficients $a(n)$ of Eq. (6) were chosen experimentally to make the quantization error minimally audible in the absence of audio with the aid of MATLAB. The coefficients obtained were: $a(1)=-0.996$, $a(2)=0.196$, $a(3)=0.112$, $a(4)=0.144$, $a(5)=-0.396$, $a(6)=-0.020$, $a(7)=0.208$, $a(8)=-0.032$, $a(9)=-0.140$, $a(10)=-0.048$, $a(11)=0.152$, and $a(12)=-0.100$.
(4) Figure 2 shows the weighted noise power N_ω of Eq. (5) (0 dB reference at the noise spectral level of no noise shaping). From this figure, it can be seen that although

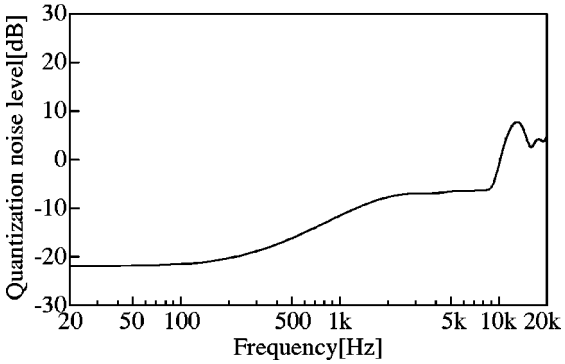


FIG. 2. Frequency response of the noise shaper transfer function.

the curve was different from our hearing threshold or equi-loudness curve, it features a reduced amount of low-frequency noise below 8.75 kHz (-6 dB), but high-frequency noise in the audible range is significantly increased. The most audible noise is concentrated at 12 kHz (6.8 dB) around the second spectral peak of the human ear's sensitivity, but the loudspeaker acts as a low-pass filter with a cutoff frequency of around 10 kHz and a slope of about -18 dB/octave.

C. Noise shaping experiment

In our experimental setup, a software version (C++ programming language) of the noise shaping requantizer was implemented with a computer (NEC: LW500J/2).

Figure 3 shows the frequency response and linearity characteristics of the noise shaper to each digital signal. Comparing Fig. 3(b) with Fig. 3(c), although both inputs are 12 bit digital signals, the results show that the quantization noise is reduced at frequencies less than 10 kHz by the noise shaping filter and that the linearity characteristic is improved to a point where it is almost equal to the characteristics for the 16 bit digital signal shown in (a). Therefore, we can see that the noise shaper acts correctly.

From the results shown in Fig. 3(b), the sound reproduction band of the loudspeaker was chosen to be from 40 Hz to 10 kHz. It is assumed that this characteristic does not significantly change even when noise shaping is applied. The quantization noise will be reduced at frequencies less than 10 kHz compared with a typical 12 bit digital signal.

III. ELECTRODYNAMIC PLANAR DIGITAL LOUDSPEAKER

A. Equivalent electric circuit model

An electrodynamic planar digital loudspeaker[1] driven by a signed magnitude binary (SMB) signal is effective. SMB is the simplest and one of the most obvious methods of encoding positive and negative numbers. The most significant bit (MSB:b_{12}) is assigned to be the sign bit. If the sign bit is 0, this means the number is positive. If the sign bit is 1, then the number is negative. The remaining 11 bits (b_{11}, b_{10}, b_9, ..., b_2, b_1) are used to represent the magnitude of the binary number in the unsigned binary notation. Therefore, the formula to convert a 12 bit input signal with a length of 12 bits is

$$V_p(t) \propto (1-2b_{12})E_0 \sum_{i=1}^{11} 2^{(i-11)} b_i(t), \tag{7}$$

where V_p is the total velocity amplitude of the panel surface, and E_0 is 16.0 V. As the diaphragm operates in the mass-controlled region, the force applied to it should, ideally, be able to operate with both polarities to reproduce the original sound pressure field correctly.

The actual active part of the diaphragm is 13 cm wide, 13 cm high, and 1.1 cm deep. The diaphragm is molded plastic (total mass $m_p = 0.0344$ kg, additional mass $m_{\mathrm{ad}} = 0.000\,88$ kg) and has streamlined sections to suppress the divided vibrations. A double-voice-coil is rolled together in one section of the diaphragm, resulting in a total of 22 voice coils, each of which has a dc impedance of 4 Ω, 22 turns, and $l = 4.142$ m. Also, 11 permanent magnets (each magnet is 7.5 mm wide, 87.25 mm long, and 6.0 mm thick, and each gap is 2.0 mm) are arranged under the diaphragm so that the poles of adjacent magnets are opposite each other. The suspension between the diaphragm and the frame is made of a piece of handmade Japanese paper (0.09 mm thick). The lowest resonant frequency f_0 was measured and found to be 22.7 Hz. Therefore, the stiffness s_p of the suspension is 736 N/m.

The free moving diaphragm generates sound in both the front and back of the element. The normal use of the element is to place it in an enclosure. In our test of the digital loudspeaker we used an enclosure ($V_c = 0.04194$ m^3) which was 28.5 cm wide, 54.5 cm high, and 27.0 cm deep. Therefore, the equivalent stiffness $s_c = \rho c^2 (ab)^2 / V_c$ is 549 N/m (20 °C).

The frequency responses of each voice coil show that at 31.0 Hz a small peak appears due to the closed enclosure and the lowest resonant frequency ($f_0 = 22.7$ Hz). From 30 to about 500 Hz, the diaphragm moves as a whole by acting as a piston. Such a loudspeaker is mass-controlled over most of its pass-band. However, when it comes to higher frequencies, the radiating area exhibits more or less strong partial vibrations. The phase does not deviate more than $\pm 35°$ in the region where the diaphragm acts as a piston,[1] but there are large phase differences, especially at frequencies larger than 500 Hz. Consequently, timing jitter and distortion may result. Therefore, our electro-dynamic planar loudspeaker has some problems that may cause large distortion when driven by a digital signal because each voice coil's response is not the same. Therefore, in order to accomplish D-A conversion at the loudspeaker, it is necessary to flatten each response. One solution is to connect in series a voice coil with a peak in its frequency response and a voice coil with a dip. We found that the acoustic responses can be improved by an appropriate combination of voice coils. Because the flatness of the frequency response from 40 Hz to 10 kHz can be improved only with the asymmetric arrangement of 11 pairs of voice coils, a 12 bit digital signal is best suited to drive the loudspeaker. We call this "the asymmetric arrangement for a 12 bit digital signal" as follows: (b_3,b_1), (b_7,b_6), (b_9,b_8), (b_5,b_4), (b_{11},b_{10}), (b_3,b_2), (b_9,b_8), (b_7,b_6), (b_{11},b_{10}), (b_5,b_4), and (b_2,b_1).

The electrical impedance will include the effect of the

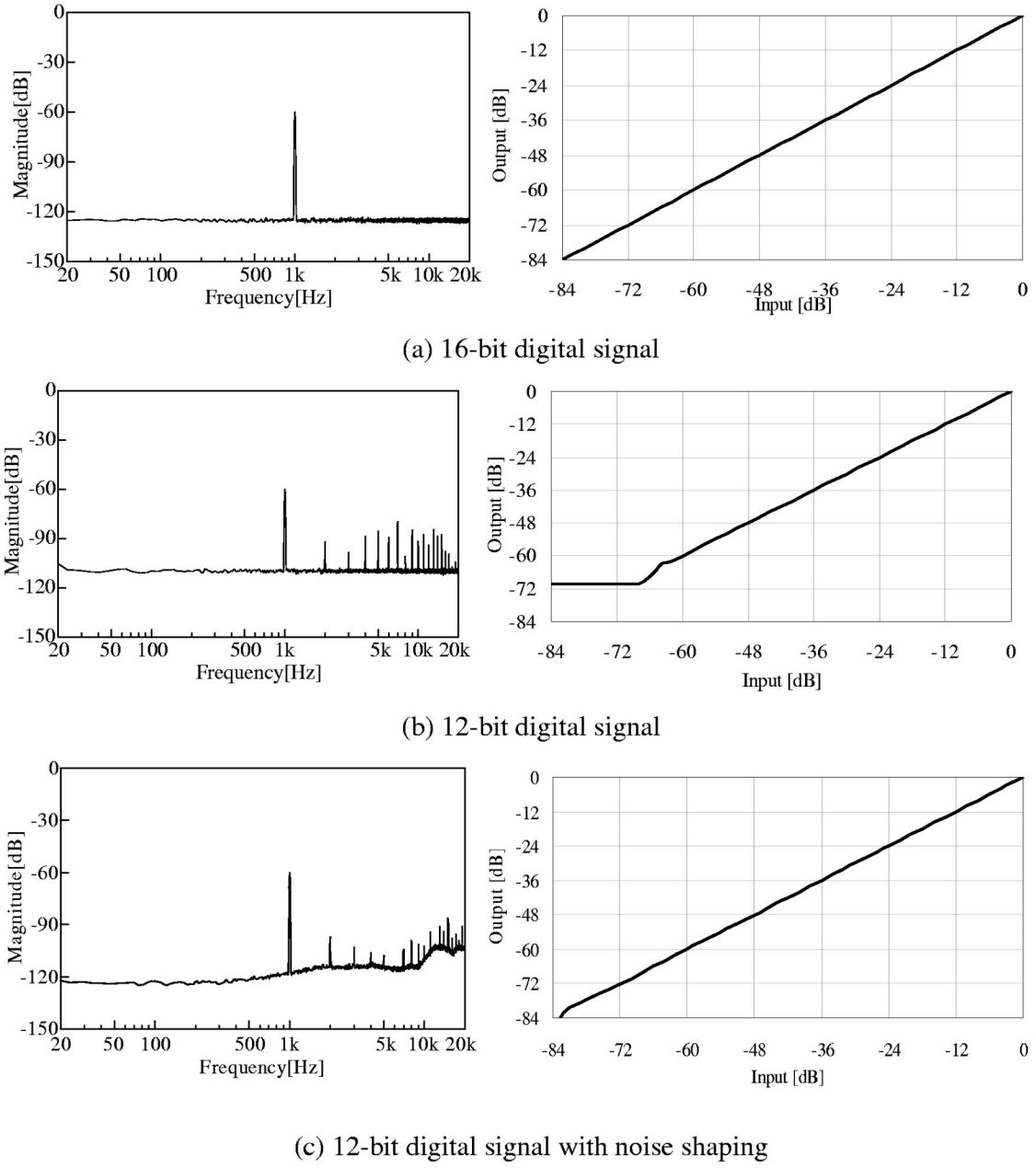


FIG. 3. The frequency response and linearity characteristics of the noise shaper when the following digital signals are input: (a) 16 bit digital signal, (b) 12 bit digital signal, and (c) 12 bit digital signal with noise shaping. The figures on the left are the frequency responses to a 1 kHz pure tone (-60 dB), while the figures on the right are the linearity characteristics at 1 kHz.

mutual coupling between each double-voice-coil: (b_3,b_1), (b_7,b_6), (b_9,b_8), (b_5,b_4), (b_{11},b_{10}), (b_3,b_2), (b_9,b_8), (b_7,b_6), (b_{11},b_{10}), (b_5,b_4), and (b_2,b_1), which are rolled together in one section of the diaphragm. Figure 4 shows a lumped-parameter equivalent circuit model[9] for a multidriver transducer consisting of a panel, a suspension, a frame and parallel exciters with magnets, and eleven double-voice-coils. In this model, R (Ω) and L (H) are the dc resistance and inductance, respectively, of each coil. M (H) represents the mutual inductance between each double-voice-coil. B (Wb/m^2) is the magnetic flux density, l (m) is the length of the ith voice coil, and Bl (N/A) is the force factor.

Consideration of each coil driven from an alternating current source of amplitude I_0 and frequency ω, and with an output impedance r_o, leads to the following series of equations:

$$r_{o3}I_{03}=(R+r_{o3}+j\omega L)I_3+j\omega MI_1+\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i,$$

$$r_{o1}I_{01}=(R+r_{o1}+j\omega L)I_1+j\omega MI_3+\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i,$$

$$r_{o7}I_{07}=(R+r_{o7}+j\omega L)I_7+j\omega MI_6+\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i,$$

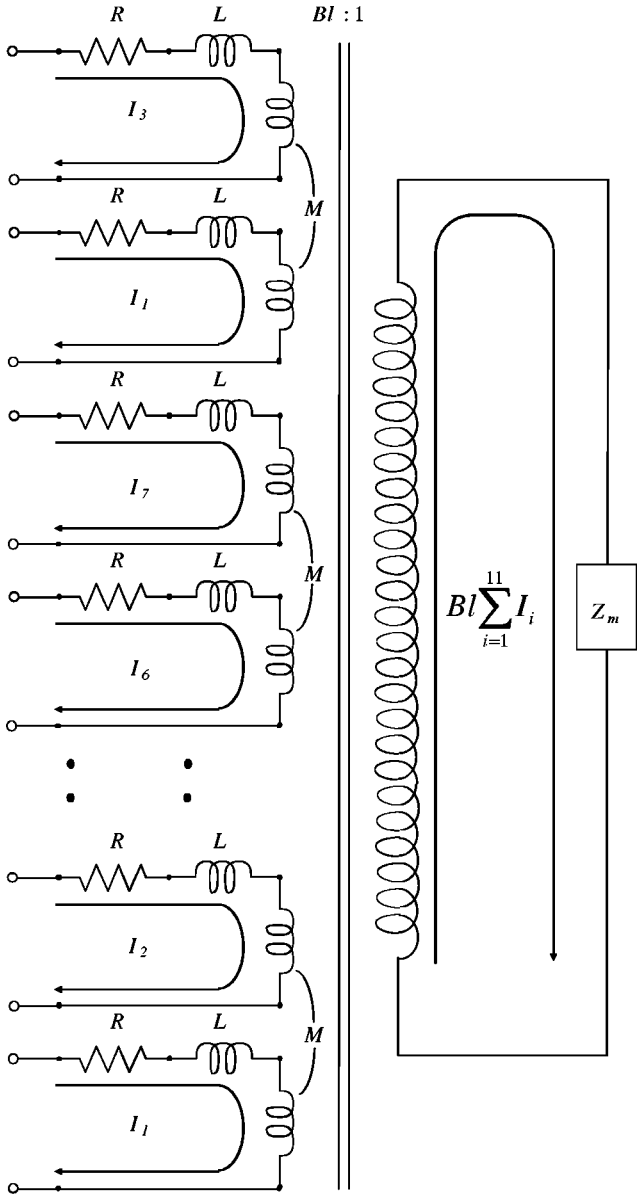


FIG. 4. Equivalent electric circuit of our electrodynamic planar digital loudspeaker.

$$
\begin{aligned}
r_{o6}I_{06} &= (R + r_{o6} + j\omega L)I_6 + j\omega M I_7 + \frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
&\vdots \\
&\vdots \\
r_{o2}I_{02} &= (R + r_{o2} + j\omega L)I_2 + j\omega M I_1 + \frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
r_{o1}I_{01} &= (R + r_{o1} + j\omega L)I_1 + j\omega M I_2 + \frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i,
\end{aligned}
\tag{8}
$$

where I_i is the current in the ith coil, and Z_m is the mechanical impedance of the motion system. The asymmetric arrangement for a 12 bit digital signal is connected in series. The voltage at the input terminals of each voice coil may be expressed as

$$
\begin{aligned}
(1-2b_{12})\times 16.0\times b_{11} &= 2(R + j\omega L)I_{11} + 2(j\omega M)I_{10} \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-1}b_{10} &= 2(R + j\omega L)I_{10} \\
&\quad + 2(j\omega M)I_{11} \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-2}b_9 &= 2(R + j\omega L)I_9 + 2(j\omega M)I_8 \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-3}b_8 &= 2(R + j\omega L)I_8 + 2(j\omega M)I_9 \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-4}b_7 &= 2(R + j\omega L)I_7 + 2(j\omega M)I_6 \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-5}b_6 &= 2(R + j\omega L)I_6 + 2(j\omega M)I_7 \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-6}b_5 &= 2(R + j\omega L)I_5 + 2(j\omega M)I_4 \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-7}b_4 &= 2(R + j\omega L)I_4 + 2(j\omega M)I_5 \\
&\quad + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-8}b_3 &= 2(R + j\omega L)I_3 + j\omega M I_1 \\
&\quad + j\omega M I_2 + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-9}b_2 &= 2(R + j\omega L)I_2 + j\omega M I_1 \\
&\quad + j\omega M I_3 + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i, \\
(1-2b_{12})\times 16.0\times 2^{-10}b_1 &= 2(R + j\omega L)I_1 + j\omega M I_2 \\
&\quad + j\omega M I_3 + 2\frac{(Bl)^2}{Z_m}\sum_{i=1}^{11} I_i,
\end{aligned}
\tag{9}
$$

where each part of Eq. (9) contains two voltage generators, V_m and V_z. $V_m = 2(j\omega M)I_i$ represents the induced electromotive force (emf) in any given coil due to the change in current in the pair coil. Likewise $V_z = 2(Bl)^2\Sigma_{i=1}^{11} I_i/Z_m$ represents the induced emf in any given coil due to the motion of the diaphragm, which is itself dependent on the total driv-

ing force on the motion system and hence on the summation of all the currents in all the voice coils.

Unfortunately the total velocity amplitude of the surface of the panel driven by 11 exciters introduces considerable complexity for two voltage generators.

The advantages of using our electrodynamic planar digital loudspeaker instead of a multiple-voice-coil digital loudspeaker[9] or conventional panel loudspeakers[10] are as follows.

(a) In a multiple-voice-coil digital loudspeaker, the motion system and the radiation impedance are the same as for a conventional loudspeaker. Since the force on the diaphragm needs to be consistent for each voice coil (when they are equally energized), they must be wound in close proximity. The electrical impedance, therefore, will include the effect of the mutual coupling between each of the voice coils. Furthermore since the current weighting in each coil must be accurately controlled, the use of a high-impedance switchable current source driver will overcome the variations in current produced by a conventional voltage source driver.

(b) A panel loudspeaker primarily consists of a panel and an inertia exciter. The exciter is essentially a voice-coil-driver with the coil attached to the panel. The magnet serves as a proof mass to produce inertial force. In lieu of a rigid diaphragm as used in conventional loudspeakers, flexible panels are employed as the primary sound radiators. Resonance of flexural motion is encouraged so that the panel vibrates as randomly as possible. The sound field produced by this type of distributed mode loudspeaker[10] is very diffuse at high frequencies. Therefore, it is difficult for conventional panel loudspeakers to achieve flatness in the frequency response from 40 Hz to 10 kHz.

B. Experiment

First, the electrical input impedance of each pair of voice coils was measured with a precision LCR meter (Agilent: 4284A, an input voltage is 100 mV). Second, when the 11 bit pair of voice coils was inputted with an 8 V_{rms} amplitude signal from 20 Hz to 20 kHz, the induced electromotive force [emf: $V_m + V_z$ (V)] for each pair of voice coils was measured with an ac voltmeter (Kenwood: VT-181E).

C. Results

The performance was measured using the following indices.

1. Input impedance

Figure 5 shows: (a) the measured input impedance $|Z_{\text{in}}|$ (Ω), and (b) the phase (deg) for the 11 and 1 bit pairs of voice coils. We see that (a) small peaks (the 11 and 1 bit pairs of voice coils) at the fundamental resonant frequency of 31.0 Hz, where $|Z_{\text{in}}| = 9.05\ \Omega$, (b) both amplitudes are about 8.8 Ω from 20 Hz to 4 kHz and both phases are less than about 10°, and (c) increases due to inductive effects at higher frequencies are from 8.8 to 12.36 Ω.

Table I shows the parameters ($2R$ and $2L$) of Eq. (9) related to the equivalent electric circuit model shown in Fig. 4. From Table I, we see that the parameter ($2R$) of each bit pair of voice coils varies from 8.89 to 10.98 Ω, and $2L$ varies from 0.0657 to 0.0731 mH.

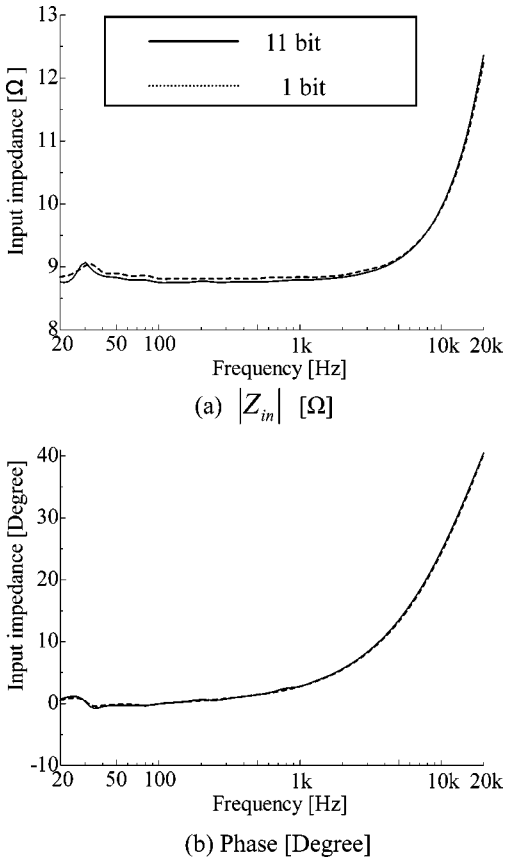


FIG. 5. Measured input impedance (amplitude and phase) of the 11 and 1 bit pairs of voice coils.

2. Induced electromotive force

Figure 6 shows the measured induced electromotive force [emf: $V_m + V_z$ (mV)] for the 10, 8, 6, and 1 bit pairs of voice coils when the 11 bit pair was driven with an 8.0 V_{rms} signal. From the lowest peak voltages of the fundamental resonant frequency, we see that the motional impedance $2\,(Bl)^2/Z_m$ of the 10 bit pair of voice coils is 0.0786 Ω, and $2\,(Bl)^2/Z_m$ of the 1 bit pair is 0.0605 Ω. Therefore, the efficiency (the ratio of the motional impedance to the free impedance of the 10 bit pair of voice coils) in converting electrical power to sound of our electrodynamic planar digital loudspeaker is about 0.87%, because the free impedance is 9.05 Ω. From the emf voltage at 10 kHz for the 10 bit pair of

TABLE I. Measured parameters of our multidriver loudspeaker.

bits	$2R$ (Ω)	$2L$ (mH)	bits	$2R$ (Ω)	$2L$ (mH)
11	9.15	0.0659	5	9.16	0.0678
10	9.02	0.0731	4	9.36	0.0707
9	10.34	0.0677	3	10.98	0.0677
8	9.03	0.0685	2	8.89	0.0657
7	9.26	0.0684	1	9.08	0.0676
6	9.07	0.0673			

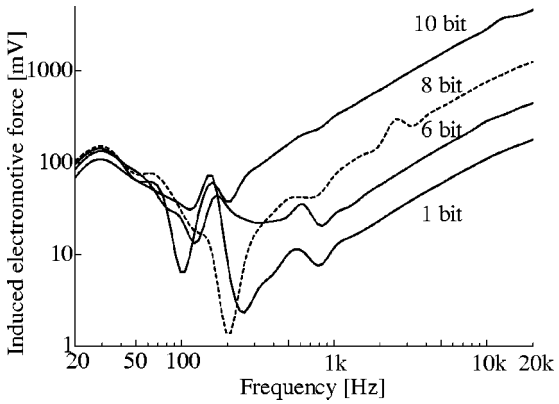


FIG. 6. Measured induced electromotive force of the 10, 8, 6, and 1 bit pairs of voice coils, when the 11 bit pair was driven with 8 V_{rms} signals.

voice coils, the mutual inductance M between each pair of coils (b_{11} and b_{10}) is 0.0287 mH, because the current of the 11 bit pair (I_{11}) is 0.791 A.

The interaction with a double-voice-coil in a multidriver loudspeaker has been modeled in terms of mutual coupling and induced motion emf. Also, our loudspeaker parameters related to the equivalent electric circuit model shown in Fig. 4 were measured: $2R=9.4\ \Omega$ on average, $2L=0.068$ mH on average, $M=0.028$ mH on average, and $2(Bl)^2/Z_m=0.074\ \Omega$ on average.

IV. RESPONSES TO 12 BIT DIGITAL SIGNALS WITH NOISE SHAPING

A D-A converter of the surface of the panel driven by 11 exciters is used to convert the sampled binary information back in to an analog vibration signal. The conversion is called a zero order hold type where each output sample level is a function of its binary weight value and is held unit the next sample arrives. As a result of the D-A converter step function response it is apparent that large amounts of undesirable high frequency energy are present. To eliminate this, the D-A converter is usually followed by a smoothing filter, having a cutoff frequency no greater than half the sampling frequency. The loudspeaker acts as a low-pass filter with a cutoff frequency of 10 kHz and a slope of about -18 dB/octave. Therefore, the spectrum of the resulting signal is the product of a step function spectrum and the band-limited analog filter spectrum. Furthermore, the radiation produced by the vibration of the surface of the rectangular panel does not have symmetric spherical radiation patterns characteristic of a simple source. However, the radiation produced by the panel can be found by considering the panel to be a group of simple point sources.[1]

A. Experiment

The experiments were carried out in an anechoic room. The on-axis response was measured at a distance of 10 cm from the front of the diaphragm.

The input signal to the loudspeaker was generated by a computer (NEC: LW500J/2), which was conveyed to the PCM driving circuit.[1] The digital audio signal had a sampling frequency of 48 kHz, and was quantized to 12 bits with noise shaping. First, a rectangular tone-burst covering a

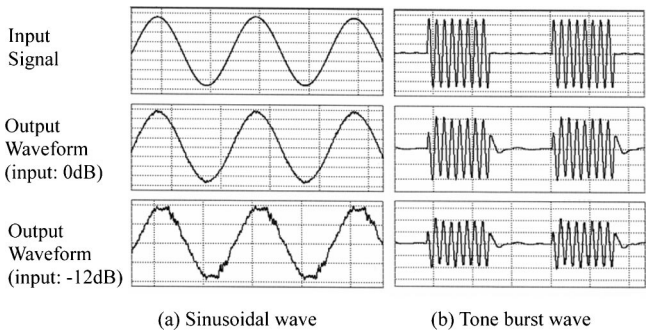


FIG. 7. Output wave forms from our electrodynamic planar digital loudspeaker driven by (a) a pure tone signal (12 bit: 0 dB, and 10 bit: -12 dB) and (b) a tone burst signal (12 bit: 0 dB, and 10 bit: -12 dB) at 100 Hz.

fairly wide frequency band was used to evaluate the dynamic behavior of the digital speaker. The transient behavior of the loudspeaker was indicated by a change in the envelope of the burst signal. Second, pure tones from 20 Hz to 20 kHz were used every 1/3 octave. The peak level of the input signal was 4096 samples (peak to peak), which corresponds to 0 dB for a 12 bit digital signal. This level was decreased in 6 dB steps and the acoustic responses were measured at each step.

B. Results

The performance was measured using the following indices.

1. Output wave forms

Example output wave forms from our electrodynamic planar digital loudspeaker driven by (a) a pure tone signal (12 bit: 0 dB, and 10 bit: -12 dB) and (b) a tone burst signal (12 bit: 0 dB, and 10 bit: -12 dB) at 100 Hz are shown in Fig. 7. From this figure, we see that our digital loudspeaker can reproduce pure sounds because the timing precision of D-A conversion on the diaphragm is good, and overhanging (ringing after the signal is gone) is present to a small degree.

2. Frequency response

Figure 8 shows the frequency responses to 12 bit digital signals (various peak levels) with noise shaping. We found that the sound output [SPL (dB)] from the asymmetric arrangement for a 12 bit digital signal with the noise shaping to input digital signal levels from 0 to -84 dB in 6 dB steps ranged from about 90 to 0 dB in 6 dB steps. We also found that the sound reproduction band was from 40 Hz to 10 kHz.

3. Linearity

For the linearity characteristics from 40 Hz to 10 kHz of the 12 bit digital signal with noise shaping, the relationship between SPL responses is shown in Fig. 8 and input levels from 0 to -84 dB were analyzed with simple regression analysis. The estimated slope coefficients are from 1.006 to 1.188, the intercept coefficients are from 84.0 to 94.0 dB, and the coefficients of determination are from 0.9972 to 0.9991. At the same time, because the linearity range is 84 dB, it can be said that the D-A conversion performed on the diaphragm is equivalent to 16 bit D-A conversion.

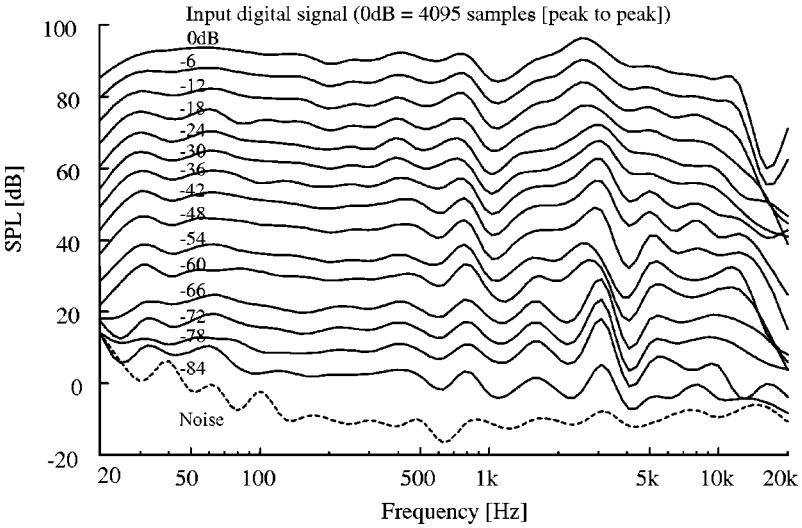


FIG. 8. Frequency response of the digital loudspeaker driven by a 12 bit digital signal with noise shaping at a distance of 10 cm from the center of the diaphragm. The dashed lines show the ambient noise level.

4. Total harmonic distortion

The total harmonic distortion (THD) of the digital loudspeaker using the asymmetric arrangement for a 12 bit digital signal is about 2% at 100 Hz as shown in Fig. 7.

Figure 9 shows the distortion factors when the input signal was changed from 0 to -12 dB in 6 dB steps. The reason why the THD is inferior at lower frequencies is that the voice coils separate from the magnetic circuit. The asymmetric arrangement for the 12 bit digital signal with noise shaping results in an improvement in the characteristics around 4 kHz, which is relatively important for the sense of hearing.

V. CONCLUSIONS

This paper extends our previous work[1] ''Acoustic characteristics of an electrodynamic planar digital loudspeaker'' by introducing a noise shaping technique into the design to minimize the quantization error resulting from the limited number of bits.

First, the noise shaping filter was designed so that (1) the number of bits used in the dithered quantizer was 12, (2) the sampling frequency was 48 kHz, and (3) the error feedback filter was a 12th order FIR design with coefficients given by: $a(1)=-0.996$, $a(2)=0.196$, $a(3)=0.112$, $a(4)=0.144$, $a(5)=-0.396$, $a(6)=-0.020$, $a(7)=0.208$, $a(8)=-0.032$, $a(9)=-0.140$, $a(10)=-0.048$, $a(11)=0.152$, and $a(12)=-0.100$.

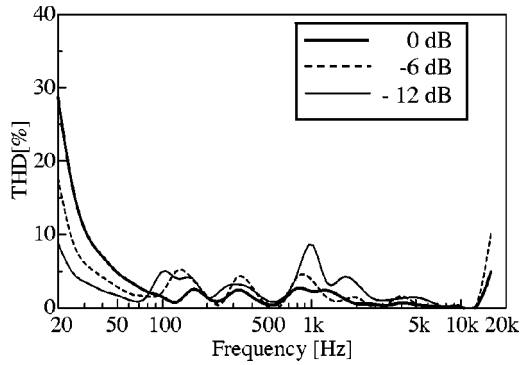


FIG. 9. Total harmonic distortion (THD) vs frequency when the input signal was changed from 0 to -12 dB in 6 dB steps.

Second, the noise shaping filter's interaction with double-voice-coils in our multidriver loudspeaker was modeled in terms of the mutual coupling and the induced motion emf. Our multidriver digital loudspeaker parameters related to the equivalent electric circuit model were measured as follows. (4) $2R=9.4\ \Omega$ on average, $2L=0.068$ mH on average, $M=0.028$ mH on average, and $2(Bl)^2/Z_m=0.074\ \Omega$ on average. (5) The efficiency (the ratio of the motional impedance to the free impedance) was about 0.87%.

Finally, to improve the linear range of our multidriver digital loudspeaker, noise shaping technology was applied to solve the problem that 12 bit PCM cannot resolve details smaller than the LSB. The responses to 12 bit digital signals with noise shaping were measured as follows. (6) Our digital loudspeaker can reproduce pure sounds because the timing precision of D-A conversion on the diaphragm is good, and overhanging (ringing after the signal is gone) is present to a small degree. (7) As for the linearity characteristics from 40 Hz to 10 kHz, the estimated slope coefficients were from 1.006 to 1.188, the intercept coefficients were from 84.0 to 94.0 dB, and the coefficients of determination were from 0.9972 to 0.9991. (8) Because the linearity range was 84 dB, it can be said that the D-A conversion performed on the diaphragm is equivalent to 16 bit D-A conversion.

To summarize, we have constructed and demonstrated a 12 bit digital loudspeaker using noise shaping. This loudspeaker has good performance, a wide bandwidth, and a linear dynamic range of 84 dB.

[1] K. Furihata, A. Hayama, D. K. Asano, and T. Yanagisawa, ''Acoustic characteristics of an electrodynamic planar digital loudspeaker,'' J. Acoust. Soc. Am. **114**, 174–184 (2003).
[2] M. A. Gerzon and P. G. Craven, ''Optimal noise shaping and dither of digital signals,'' Audio Engineering Society 87th Convention, New York, 1989, preprint, p. 2822.
[3] M. A. Gerzon, P. G. Craven, J. R. Stuart, and R. J. Wilson, ''Psychoacoustic noise shaped improvements in CD and other linear digital media,'' Audio Engineering Society 94th Convention, Berlin, 1993, reprint, p. 3501.
[4] J. R. Stuart and R. J. Wilson, ''A search for efficient dither for DSP applications,'' On the possibility and problems of the PCM digital loud-

speaker (in Japanese), Audio Engineering Society 92nd Convention, Vienna, 1992, preprint, p. 3334.
[5] M. Akune, R. M. Heddle, and K. Akagiri, ''Super Bit Mapping: Psychoacoustically optimized digital recording,'' Audio Engineering Society 93rd Convention, San Francisco, 1992, preprint, p. 3371.
[6] W. Verhelst and D. D. Koning, ''Least squares theory and design of optimal noise shaping filters,'' Audio Engineering Society 22nd International Conference on Virtual, Synthetic and Entertainment Audio, Espoo, Finland, 2002, preprint, pp. 216–222.
[7] S. P. Lipshitz, J. Vanderkooy, and R. A. Wannamaker, ''Minimally audible noise shaping,'' J. Audio Eng. Soc. **39**, 836–852 (1991).
[8] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech* (Springer, New York, 1976).
[9] Y. Huang, S. C. Busbridge, and P. A. Fryer, ''Interactions in a multiple-voice-coil digital loudspeaker,'' J. Audio Eng. Soc. **48**, 545–552 (2000).
[10] M. R. Bal and T. Huang, ''Development of panel loudspeaker system: Design, evaluation and enhancement,'' J. Acoust. Soc. Am. **109**, 2751–2761 (2001).

Forced responses of solid axially polarized piezoelectric ceramic finite cylinders with internal losses

D. D. Ebenezer,[a] K. Ravichandran, and R. Ramesh
Naval Physical and Oceanographic Laboratory, Thrikkakara, Kochi 682 021, India

Chandramouli Padmanabhan
Department of Mechanical Engineering, Indian Institute of Technology—Madras, Chennai 600 036, India

(Received 23 September 2004; revised 2 February 2005; accepted 1 March 2005)

A method is presented to determine the forced responses of piezoelectric cylinders using weighted sums of only certain exact solutions to the equations of motion and the Gauss electrostatic conditions. One infinite set of solutions is chosen such that each field variable is expressed in terms of Bessel functions that form a complete set in the radial direction. Another infinite set of solutions is chosen such that each field variable is expressed in terms of trigonometric functions that form a complete set in the axial direction. Another solution is used to account for the electric field that can exist even when there is no vibration. The weights are determined by using the orthogonal properties of the functions and are used to satisfy specified, arbitrary, axisymmetric boundary conditions on all the surfaces. Special cases including simultaneous mechanical and electrical excitation of cylinders are presented. All numerical results are in excellent agreement with those obtained using the finite element software ATILA. For example, the five lowest frequencies at which the conductance and susceptance of a stress-free cylinder, of length 10 mm and radius 5 mm, reach a local maximum or minimum differ by less than 0.01% from those computed using ATILA. [DOI: 10.1121/1.1900543]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.20.Bi, 43.40.At [AJZ] Pages: 3645–3656

I. INTRODUCTION

Free vibration analysis is simpler than forced vibration analysis and is done more often because the knowledge of natural frequencies and eigenfunctions is sometimes sufficient. When the forced response of lossless structures is of interest, it is sometimes determined by using the eigenfunctions that form a complete set of functions. However, eigenfunctions are often expressed in terms of infinite series or determined using numerical methods and it may, therefore, not be convenient to use them to determine the forced response. Closed form expressions for eigenfunctions are known only for some simple structures with simple boundary conditions. Further, eigenfunctions or normal modes do not generally exist for structures with internal losses such as piezoelectric cylinders. In this paper, complete sets of orthogonal functions are used to determine the forced responses of solid, axially polarized, piezoelectric, ceramic cylinders with internal losses. The functions are exact solutions to the governing equations.

Complete sets of functions have been used in conjunction with several methods to study the vibration of a variety of structures and geometries. Chung[1] used Fourier series for free vibration analysis of finite, elastic, cylindrical shells while Wang and Lin[2] used it to investigate several structural mechanics problems. The Fourier series has also been used by Hurlebaus *et al.*[3] to determine the resonance frequencies and eigenfunctions of free, elastic, orthotropic plates; by Bisegna and Maceri[4] to analyze simply supported rectangular piezoelectric plates; and by Kapuria *et al.*[5] to analyze simply supported piezoelectric cylindrical shells. Ding *et al.*[6] used the Fourier series to analyze piezoelectric panels and Sharma[7] used it for the analysis of piezothermoelastic panels. The Fourier-Bessel series has been used by Holland and EerNisse[8] to analyze finite, solid, piezoelectric cylinders; by Hutchinson[9] to analyze the free vibration of finite, solid, elastic cylinders; and by Junger[10] to study the sound field inside a finite, fluid cylinder with specified nonuniform displacements at the boundaries. The investigations using complete sets of Fourier-Bessel functions fall into two categories. In the first category, each term in the series is an exact solution to the equations of motion. Therefore, the truncated series is also an exact solution to the equations of motion and only the boundary conditions are approximately satisfied. In the second category, both the equations of motion and the boundary conditions are approximately satisfied. Investigations by Hutchinson[9] and Junger[10] belong to the first category while the work of Holland and EerNisse[8] is in the second category.

Recently, Huang *et al.*[11] used closed-form expressions to analyze the free vibration of axially polarized, short-circuited, piezoelectric cylinders. They compared their computed resonance frequencies with the experimental results of Shaw[12] and found that even the edge mode resonance frequencies differ by less than 3%. EerNisse[13] used a variational approach to compute the resonance frequencies of axially polarized cylinders and obtained agreement with Shaw[12] of about 3% for the lowest eight branches. The computed resonance frequencies had a convergence of better than 1%.

This paper is an extension of the investigation of the forced vibration of piezoelectric and elastic solid cylinders. Ebenezer and Ramesh[14] investigated the forced response of

[a]Electronic mail: tsonpol@vsnl.com

axially polarized cylinders with electrodes on only the flat surfaces. They determined the exact response of cylinders with arbitrary length-to-radius ratio to certain uniform excitations. The closed-form expressions they used for the axial and radial components of displacement and potential are the first terms in expansions of these functions in terms of complete sets of functions. Prasad and Jain,[15] who studied anisotropic finite cylinders, and Paul,[16] who studied piezoelectric finite cylinders, did not consider these leading terms. Ramesh and Ebenezer[17] also showed that these terms are sufficient to approximately determine the response of stress-free axially polarized rings in the neighborhoods of the first resonance frequency in the lower and upper branches. The rings are often used in piezoelectric sonar transducers. Their wall thickness and length are comparable but much less than the diameter. Ebenezer and Ramesh[18] also analyzed axially polarized cylinders with arbitrary boundary conditions on the flat surfaces using the Bessel series. Only the leading term in the Fourier series was used in the expression for the radial displacement and only certain uniform boundary conditions could be satisfied on the curved surface. Ebenezer *et al.*[19] used the same approach used in this paper to analyze the free and forced vibration of elastic cylinders. The resonance frequencies of free elastic cylinders, including the edge mode resonance frequencies, computed by them differ from those obtained using the finite element method by less than 0.1%. The mechanical and electrical forced responses[14,17–19] are also in good agreement with those obtained using ATILA[20]—a finite element package for the analysis of sonar transducers.

In this paper, forced vibrations of solid, axially polarized, piezoelectric, ceramic, finite cylinders with arbitrary length-to-radius ratio, electrodes on only the top and bottom flat surfaces, and internal losses are analyzed using exact, linearized, governing equations. Losses are modeled using complex dielectric, elastic, and piezoelectric coefficients. The response to arbitrary, piecewise continuous, and localized excitations acting on the flat and curved surfaces is determined using weighted sums of certain exact solutions to the equations of motion and the Gauss electrostatic conditions. Axial and radial components of displacement and potential are used as primary variables. Stress and electric displacement are expressed in terms of the weighted functions used to describe the primary variables. One infinite set of solutions is chosen such that each field variable is expressed in terms of Bessel functions that form a complete and orthogonal set in the radial direction. Another infinite set of solutions is chosen such that each field variable is expressed in terms of trigonometric functions that form a complete and orthogonal set in the axial direction. Another solution is chosen to account for the electric field that can exist even when there is no vibration. Arbitrary boundary conditions can be satisfied on the surfaces because the sets are complete. The weights are easily determined by using the orthogonal property to satisfy the boundary conditions. Special cases including simultaneous mechanical excitation over a portion of the cylinder and electrical excitation are presented. Numerical results are also presented to illustrate the response of cylinders to various excitations and are shown to converge very rapidly to those obtained using finite element analysis.

Internal losses give rise to hysteresis loops and internal heating. The effect of internal losses can be modeled using complex coefficients and the linearized governing equations under steady-state conditions. However, when transients are of interest, this approach is not suitable and nonlinear analysis is essential.

The single cylinder that is analyzed here is a building block in a number of devices. For example, Langevin transducers have piezoelectric cylinders sandwiched between elastic cylinders and 1–3 piezocomposites have solid piezoelectric cylinders embedded in hollow elastic cylinders. Therefore, the analysis of cylinders with arbitrary boundary conditions can be extended to analyze such devices. For example, the classical Langevin transducer that is approximately analyzed by Iula *et al.*[21] can be exactly analyzed using the models of the solid elastic cylinder in Ref. 19 and solid piezoelectric cylinder presented here.

II. THEORETICAL ANALYSIS

Consider a solid, axially polarized, piezoelectric, ceramic, finite cylinder of length L and radius a as shown in Fig. 1. The top and bottom flat surfaces are fully electroded. Arbitrary but axisymmetric boundary conditions are specified on the flat and curved surfaces. The frequency-dependent response of the cylinder is of interest. It can be determined for non-axisymmetric boundary conditions[22] by using Fourier series expansions in the tangential direction.

The excitation and, therefore, the response of the cylinder are axisymmetric. The exact equilibrium equations in cylindrical coordinates (r,θ,z), the expressions for strain in terms of displacement and for electric field in terms of potential, the constitutive relations for a piezoelectric material of 6*mm* crystal class, the Gauss electrostatic condition, and the exact axisymmetric equations of motion used here are the same as those in Eqs. (1)–(5) in Ref. 18. In these Eqs. U and W are the axial and radial displacements respectively, T_{rr}, T_{zz} and $T_{\theta\theta}$ are normal components of stress, T_{rz} is a shear component of stress, Φ is the electric potential, D_z and

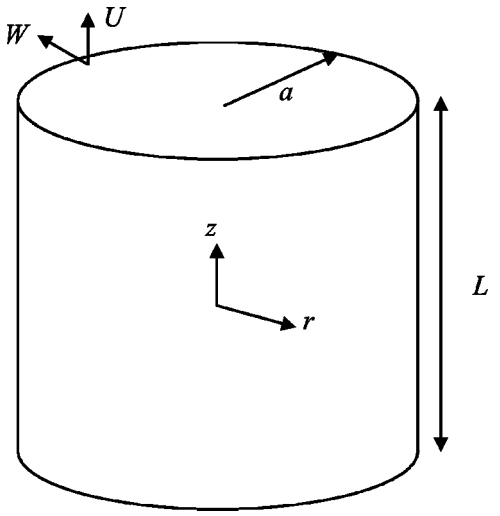


FIG. 1. A solid axially polarized cylinder of length L and radius a. The axial and radial components of displacement are U and W, respectively.

D_r are components of electric displacement, ρ is the density, ω is the angular frequency, c_{11}^E, c_{12}^E, c_{13}^E, c_{33}^E, and c_{44}^E are the eleastic stiffness coefficients, e_{15}, e_{31}, and e_{33} are the piezoelectric coeficients, and ε_{11}^S and ε_{33}^S are the dielectric permittivities. Further, time-harmonic variation of variable is assumed, extensional stress is positive, and $e^{j\omega t}$ is omitted for convenience.

It is easily verified that

$$[U\ \ W\ \ \Phi]^T=[U_1\ \ W_1\ \ \Phi_1]^T+[U_2\ \ W_2\ \ \Phi_2]^T+[U_3\ \ W_3\ \ \Phi_3]^T \tag{1}$$

is the sum of three exact solutions to Eq. (5) of Ref. 18 where

$$\begin{Bmatrix} U_1 \\ W_1 \\ \Phi_1 \end{Bmatrix} = \begin{Bmatrix} A\sin(K_{z0}z) \\ BJ_1(k_{r0}r) \\ A\dfrac{e_{33}}{\varepsilon_{33}^S}\sin(K_{z0}z)+Dz+E \end{Bmatrix}, \tag{2a}$$

$$K_{z0}=\omega/\nu^D, \quad \nu^D=\left(\frac{c_{33}^E+e_{33}^2/\varepsilon_{33}^S}{\rho}\right)^{1/2}=\left(\frac{c_{33}^D}{\rho}\right)^{1/2},$$

$$k_{r0}=\omega/\nu^E, \quad \nu^E=\left(\frac{c_{11}^E}{\rho}\right)^{1/2}, \tag{2b}$$

$$\begin{Bmatrix} U_2 \\ W_2 \\ \Phi_2 \end{Bmatrix} = \begin{Bmatrix} \sum_{m=1}^{M_r}\sum_{s=1}^{3} A_{ms}J_0(k_{rm}r)\sin(k_{zms}z) \\ \sum_{m=1}^{M_r}\sum_{s=1}^{3} A_{ms}\psi_{ms}J_1(k_{rm}r)\cos(k_{zms}z) \\ \sum_{m=1}^{M_r}\sum_{s=1}^{3} A_{ms}\chi_{ms}J_0(k_{rm}r)\sin(k_{zms}z) \end{Bmatrix}, \tag{2c}$$

$$\begin{Bmatrix} U_3 \\ W_3 \\ \Phi_3 \end{Bmatrix} = \begin{Bmatrix} \sum_{m=1}^{M_r}\sum_{s=1}^{3} B_{ms}J_0(K_{rms}r)\sin(K_{zm}z) \\ \sum_{m=1}^{M_r}\sum_{s=1}^{3} B_{ms}\phi_{ms}J_1(K_{rms}r)\cos(K_{zm}z) \\ \sum_{m=1}^{M_r}\sum_{s=1}^{3} B_{ms}\gamma_{ms}J_0(K_{rms}r)\sin(K_{zm}z) \end{Bmatrix}. \tag{2d}$$

and the weights A, B, A_{ms}, and B_{ms} in Eq. (2) are used to satisfy arbitrary boundary conditions on all the surfaces. The first exact solution in Eq. (1) is used in Refs. 14 and 17. ν^D, in Eq. (2b), is also associated with thickness mode waves[23] in thin piezoelectric plates with electric field parallel to thickness. ν^E, in Eq. (2b), is apparently used for the first time in Ref. 14. The first two exact solutions in Eq. (1) are used in Ref. 18. The third solution makes it possible to satisfy arbitrary boundary conditions on the curved surface.

In the above equations, J_μ is the μth-order Bessel function of the first kind. Equations (2c) and (2d) are made exact solutions for arbitrary values of k_{rm}, $m=1,2,3,...M_r$, and K_{zm}, $m=1,2,3,...M_z$, by suitably choosing the frequency-dependent values of k_{zms}, K_{rms}, χ_{ms}, ψ_{ms}, γ_{ms}, and ϕ_{ms}.

The frequency-dependent values of k_{zms} are determined by substituting Eq. (2c) in Eq. (5) of Ref. 18 and equating the determinant of the resulting equation to zero. The characteristic equation is cubic in k_{zms}^2 and is solved for $m=1,2,...,M_r$. Similarly, the frequency-dependent values of K_{rms} are determined by substituting Eq. (7d) in Eq. (5) of Ref. 18 and equating the determinant of the resulting equation to zero. The characteristic equation is cubic in K_{rms}^2 and solved for $m=1,2,...,M_z$. The corresponding equations are quadratic[19] in isotropic elastic materials and simple expressions are obtained for k_{zms} and K_{rms}.

Substituting Eq. (2c) in Eq. (5) of Ref. 18 and rearranging terms yields the following expressions for ψ_{ms} and χ_{ms}:

$$\psi_{ms}=\{[\rho\omega^2-c_{33}k_{zms}^2-c_{44}^Ek_{rm}^2][\varepsilon_{33}^Sk_{zms}^2+\varepsilon_{11}^Sk_{rm}^2]-[e_{33}k_{zms}^2+e_{15}k_{rm}^2]^2\}/\Delta_k, \tag{3a}$$

and

$$\chi_{ms}=\{[\rho\omega^2-c_{33}k_{zms}^2-c_{44}^Ek_{rm}^2][(e_{15}+e_{31})k_{rm}k_{zms}]+[e_{33}k_{zms}^2+e_{15}k_{rm}^2][(c_{13}^E+c_{44}^E)k_{rm}k_{zms}]\}/\Delta_k, \tag{3b}$$

where

$$\Delta_k=[e_{33}k_{zms}^2+e_{15}k_{rm}^2][(e_{15}+e_{31})k_{rm}k_{zms}]+[\varepsilon_{33}^Sk_{zms}^2+\varepsilon_{11}^Sk_{rm}^2][(c_{13}^E+c_{44}^E)k_{rm}k_{zms}]. \tag{3c}$$

Similarly, substituting Eq. (2d) in Eq. (5) of Ref. 18 and rearranging terms yields the following expressions for ϕ_{ms} and γ_{ms}:

$$\phi_{ms}=\{[\rho\omega^2-c_{44}^EK_{rms}^2-c_{33}^EK_{zm}^2][\varepsilon_{11}^SK_{rms}^2+\varepsilon_{33}^SK_{zm}^2]-[e_{15}K_{rms}^2+e_{33}K_{zm}^2]^2\}/\Delta_K, \tag{4a}$$

and

$$\gamma_{ms}=\{[\rho\omega^2-c_{44}^EK_{rms}^2-c_{33}^EK_{zm}^2][(e_{15}+e_{31})K_{zm}K_{rms}]+[e_{15}K_{rms}^2+e_{33}K_{zm}^2][(c_{13}^E+c_{44}^E)K_{zm}K_{rms}]\}/\Delta_K, \tag{4b}$$

where

$$\Delta_K=[e_{15}K_{rms}^2+e_{33}K_{zm}^2][(e_{15}+e_{31})K_{zm}K_{rms}]+[\varepsilon_{11}^SK_{rms}^2+\varepsilon_{33}^SK_{zm}^2][(c_{13}^E+c_{44}^E)K_{zm}K_{rms}]. \tag{4c}$$

Here, $k_{rm}a$ are chosen to be the roots of $J_1(k_{rm}a)=0$ and are approximately equal to 0, 3.83, 7.02, 10.17,... for $m=0,1,2,3,...$, respectively. When $m=0$, $k_{rm}a$ and the radial displacement are zero but the axial displacement and potential are nonzero and are functions of K_{z0}. Therefore, for convenience, the $m=0$ term is not included in Eq. (2c) but in Eq. (2a). Similarly, $K_{zm}L/2$ are chosen to be $m\pi$ with $m=0,1,2,3,...$, respectively. When $m=0$, the axial displacement and potential are zero and the radial displacement is a function of K_{r0}. Therefore, for convenience, the $m=0$ term is not included in Eq. (2d) but in Eq. (2a).

In addition to the leading terms of the Bessel and Fourier series, Eq. (2a) contains two terms for the potential with coefficients D and E. These terms are required because a linear variation of potential can exist even when the displacement is zero due to inertial clamping or clamped boundary conditions.[8]

The values of $k_{rm}a$ and $K_{zm}L/2$ are chosen such that the field variables are expressed in terms of complete sets of functions on the flat and curved surfaces, respectively. It is noted that for $M_r=\infty$, $J_\mu(k_{rm}r)$ form a point-wise complete set of orthogonal functions when $\mu=0$ and a norm-wise complete set of orthogonal functions when $\mu=1$[18,24]. (Sets of functions that are not all zero at the same point and form a norm-wise complete set are known as point-wise complete sets.) Similarly, for $M_z=\infty$, $\sin(K_{zm}z)$ and $\cos(K_{zm}z)$ form norm-wise and point-wise complete sets of orthogonal functions, respectively, i.e.,

$$\int_{-L/2}^{+L/2} H(K_{zm}z)H(K_{zn}z)dz=\begin{cases} 0, & m\neq n, \\ L/2, & m=n, \end{cases} \tag{5}$$

where $H(.)$ is either sin(.) or cos(.)

The origin of the coordinate system is fixed at the center of the cylinder and the excitation is symmetric about the center. Therefore, the radial displacement will be symmetric about the center of the cylinder and it is expressed in terms of cos functions in Eq. (2). If the excitation is antisymmetric about the center, then the radial displacement will also be antisymmetric and it should be in terms of sin functions. Corresponding changes should also be made in the expressions for axial displacement and potential. The Bessel function of the second kind has not been included in Eqs. (2c) and (2d) in order to satisfy the finiteness condition at the origin. If the cylinder is hollow, this function should also be included.

Other quantities of interest are now easily determined by using Eqs. (2), (3), of Ref. 18 and (1). For example, the normal components of stress and charge density in the axial direction are expressed as

$$\begin{aligned} T_{zz}=&A\left(c_{33}^E+\frac{e_{33}^2}{\varepsilon_{33}^S}\right)K_{z0}\cos(K_{z0}z)+Bc_{13}^E k_{r0}J_0(k_{r0}r) \\ &+De_{33}+\sum_{m=1}^{M_r}\sum_{s=1}^{3}A_{ms}[c_{13}^E\psi_{ms}k_{rm}+(c_{33}^E \\ &+e_{33}\chi_{ms})k_{zms}]J_0(k_{rm}r)\cos(k_{zms}z) \\ &+\sum_{m=1}^{M_z}\sum_{s=1}^{3}B_{ms}[c_{13}^E\phi_{ms}K_{rms}+(c_{33}^E \\ &+e_{33}\gamma_{ms})K_{zm}]J_0(K_{rms}r)\cos(K_{zm}z). \end{aligned} \tag{6}$$

and

$$\begin{aligned} D_z=&Be_{31}k_{r0}J_0(k_{r0}r)-D\varepsilon_{33}^S+\sum_{m=1}^{M_r}\sum_{s=1}^{3}A_{ms}[e_{31}\psi_{ms}k_{rm} \\ &+(e_{33}-\varepsilon_{33}^S\chi_{ms})k_{zms}]J_0(k_{rm}r)\cos(k_{zms}z) \\ &+\sum_{m=1}^{M_z}\sum_{s=1}^{3}B_{ms}[e_{31}\phi_{ms}K_{rms}+(e_{33} \\ &-\varepsilon_{33}^S\gamma_{ms})K_{zm}]J_0(K_{rms}r)\cos(K_{zm}z). \end{aligned} \tag{7}$$

respectively, and the current is expressed as

$$\begin{aligned} I=&-\int_0^a j\omega D_z 2\pi r\,dr=-j\omega\pi a(2Be_{31}J_1(k_{r0}a) \\ &-D\varepsilon_{33}^S a)-j2\pi\omega\sum_{m=1}^{M_z}\cos(K_{zm}z) \\ &\times\left\{\sum_{s=1}^{3}B_{ms}[e_{31}\phi_{ms}K_{rms}+(e_{33} \right. \\ &\left.-\varepsilon_{33}^S\gamma_{ms})K_{zm}]\frac{a}{K_{rms}}J_1(K_{rms}a)\right\} \end{aligned} \tag{8}$$

because[24]

$$\int_0^a J_0(k_{rm}r)r\,dr=0, \quad m=1,2,3,\ldots. \tag{9}$$

The current in Eq. (8) should be independent of z, as seen from the Gauss zero-divergence condition of Eq. (4) in Ref. 18. It, therefore, follows that the term in braces, i.e., in { }, should be zero. This is seen to be the case for each of the special cases considered below. It is noted that the current depends only on the average boundary conditions, i.e., it is independent of A_{ms} and B_{ms}. The input electrical admittance,

$$Y=G+jB=I/\Phi_0 \tag{10}$$

is easily obtained by using Eq. (8) and the specified applied voltage, Φ_0, where G and B are the input electrical conductance and susceptance, respectively. The real part of the input current is positive when the cylinder is electrically excited but may be negative when the cylinder is simultaneously electrically and mechanically excited.

It is seen from the above that arbitrary boundary conditions can be satisfied because all field variables are expressed in terms of complete sets of functions. On the flat, electroded surfaces, U, Φ, and T_{zz} are expressed in terms of $J_0(k_{rm}r)$, $m=0,1,2,...$; and W and T_{rz} are expressed in terms of $J_1(k_{rm}r)$, $m=1,2,\ldots$. On the curved, unelectroded surfaces, U, Φ, and T_{rz} are expressed in terms of $\sin(K_{zm}z)$, $m=1,2,...$; and W and T_{rr} are expressed in terms of $\cos(K_{zm}z)$, $m=0,1,2,\ldots$. It therefore follows that arbitrary, piecewise continuous, and localized boundary and continuity conditions can be satisfied.

Special cases

Several special cases are considered now. The stress on the surface of the cylinder and the electrical boundary conditions are specified in all the cases. However, a similar procedure can be used when the displacements are specified.

The normal and shear stress on the flat surfaces are denoted by $\bar{T}_{zz}$ and $\bar{T}_{rz}$, respectively, and the normal and shear stress on the curved surface are denoted by $\hat{T}_{rr}$ and $\hat{T}_{rz}$, respectively. The potential on the bottom electrode is zero and the potential on the top electrode is uniform, i.e., $\Phi=\Phi_0$ on the top flat surface. Φ_0 is unity for the applied potential condition, zero for the short-circuit condition, and is to be determined for the open-circuit condition. The normal components of electric displacement on the flat and

TABLE I. Material properties[25] of PZT5A.

Material property	Value
ρ (kg/m^3)	7750
ε_{11}^S (pF/m)	916×8.854
ε_{33}^S	830×8.854
c_{11}^E (GN/m^2)	121
c_{12}^E	754
c_{13}^E	752
c_{33}^E	111
c_{44}^E	21.1
e_{31} (C/m^2)	−5.4
e_{33}	15.8
e_{15}	12.3
v^D (m/s)	4325
v^E	3951

curved surface are denoted by $\overline{D}_z$ and $\hat{D}_r$, respectively. Under open-circuit condictions, the current is zero. Therefore,

$$\int_0^a \overline{D}_z 2\pi r\, dr = 0, \tag{11}$$

The boundary condition on T_{zz} is satisfied by using the orthogonal property of $J_0(k_{rn}r)$ in Eq. (7) of Ref. 18. Substituting the specified function T_{zz} in Eq. (6), multiplying both sides by $rJ_0(k_{rn}r)$, and integrating over r yields

$$\begin{aligned} &A\left[\left(c_{33}^E + \frac{e_{33}^2}{\varepsilon_{33}^s}\right)\frac{K_{z0}a^2}{2}\cos\left(\frac{K_{z0}L}{2}\right)\right] + B[c_{13}^E a J_1(K_{r0}a)] \\ &\quad + D_z\left[e_{33}\frac{a^2}{2}\right] + \sum_{m=1}^{M_z}\sum_{s=1}^{3} B_{ms}[c_{13}^E \phi_{ms} K_{rms} + (c_{33}^E \\ &\quad + e_{33}\gamma_{ms})K_{zm}](-1)^m \frac{a}{K_{rms}} J_1(K_{rms}a) \\ &\quad = \int_0^a \overline{T}_{zz} r\, dr, \quad n = 0 \end{aligned} \tag{12a}$$

and

$$\begin{aligned} &B[c_{13}^E K_{r0} R(K_{r0})] + \sum_{s=1}^{3} A_{ns}[c_{13}^E \psi_{ns} k_{rn} + (c_{33}^E \\ &\quad + e_{33}\chi_{ns})k_{zns}]R(k_{rn})\cos\left(\frac{k_{zns}L}{2}\right) \\ &\quad + \sum_{m=1}^{M_z}\sum_{s=1}^{3} B_{ms}[c_{13}^E \phi_{ms} K_{rms} \\ &\quad + (c_{33}^E + e_{33}\gamma_{ms})K_{zm}](-1)^m R(K_{rms}) \\ &\quad = \int_0^a \overline{T}_{zz} r J_0(k_m r)\, dr, \quad n = 1,2,\ldots. \end{aligned} \tag{12b}$$

The above equation is obtained by using[24]

$$\int_0^a r J_0(Xr) J_0(k_{rn}r)\, dr = R(X), \tag{13a}$$

where

$$R(X) = \begin{cases} 0, & X = k_{rm},\ m \neq n \\ \dfrac{a^2}{2} J_0^2(k_{rn}a), & X = k_{rn},\ m = n \\ \dfrac{Xa}{X^2 - k_{rn}^2} J_0(k_{rn}a) J_1(Xa), & X \neq k_{rn},\ n = 1,2,3,\ldots \end{cases}. \tag{13b}$$

The boundary condition on T_{rz} on the flat surfaces, is satisfied by using the orthogonal property of of $J_1(k_{rn}r)$ in Eq. (7) of Ref. 18. Substituting the specified function T_{rz} in the general expression for T_{rz}, multiplying both sides by $rJ_1(k_{rn}r)$, and integrating over r yields

$$\begin{aligned} &-\frac{a^2}{2} J_0^2(k_{rn}a)\sum_{s=1}^{3} A_{ns}[(c_{44}^E + e_{15}\chi_{ns})k_{rn} \\ &\quad + \psi_{ns}k_{zns}c_{44}^E]\sin(k_{zns}L/2) \\ &\quad = \int_0^a \overline{T}_{rz} r J_1(k_{rn}r)\, dr, \quad n = 1,2,\ldots. \end{aligned} \tag{14}$$

The boundary condition on T_{rr} on the flat surfaces, is satisfied by using the orthogonal property of $\cos(K_{zm}z)$ in Eq. (5). Substituting the specified function $\hat{T}_{rr}$ in the general expression for T_{rr}, multiplying both sides by $\cos(K_{zn}z)$, and integrating over z yields

$$\begin{aligned} &A\left[2\left(c_{13}^E + \frac{e_{31}e_{33}}{\varepsilon_{33}^s}\right)\sin\left(\frac{K_{z0}L}{2}\right)\right] + B\left[c_{11}^E K_{r0} J_0(K_{r0}a) \right. \\ &\quad \left. + \frac{(c_{12}^E - c_{11}^E)}{a} J_1(K_{r0}a)\right]L + D[e_{31}L] \\ &\quad + \sum_{m=1}^{M_r}\sum_{s=1}^{3} A_{ms} 2\left[\frac{c_{11}^E}{k_{zms}}\psi_{ms}k_{rm} + c_{13}^E \right. \\ &\quad \left. + e_{31}\chi_{ms}\right] J_0(k_{rm}a)\sin\left(\frac{k_{zms}L}{2}\right) \\ &\quad = \int_{-L/2}^{L/2} \hat{T}_{rr}\, dz, \quad n = 0, \end{aligned} \tag{15a}$$

and

$$\begin{aligned} &A\left(c_{13}^E + \frac{e_{31}e_{33}}{\varepsilon_{33}^s}\right)K_{z0}S(K_{z0}) + \sum_{m=1}^{M_r}\sum_{s=1}^{3} A_{ms}[c_{11}^E \psi_{ms}k_{rm} \\ &\quad + (c_{13}^E + e_{31}\chi_{ms})k_{zms}]J_0(k_{rm}a)S(k_{zms}) \\ &\quad + \sum_{s=1}^{3} B_{ns}\left\{[c_{11}^E \phi_{ns}K_{rns} + (c_{13}^E \right. \\ &\quad \left. + e_{31}\gamma_{ns})K_{zn}]J_0(K_{rns}a) + \frac{\phi_{ns}}{a} J_1(K_{rns}a)(c_{12}^E \right. \\ &\quad \left. - c_{11}^E)\right\} S(K_{zn}) \\ &\quad = \int_{-L/2}^{L/2} \hat{T}_{rr}\cos(K_{zn}z)\, dz, \quad n = 1,2,\ldots. \end{aligned} \tag{15b}$$

TABLE II. Resonance frequencies of a solid lossless PZT5A cylinder of radius 20.05 mm and height 2.03 mm computed using various methods.

Mode no.	Resonance frequency (kHz)				
	Guo and Cawley[25]	ATILA[20] $I=J=40$	Present method		
			$M_r=20,\ M_z=2$	$M_r=50,\ M_z=5$	$M_r=100,\ M_z=10$
1	49.56	49.56	49.56	49.56	49.56
2	128.1	128.1	128.1	128.1	128.1
3	201.6	201.6	201.6	201.6	201.6
4	272.1	272.0	272.1	272.1	272.0
5	338.5	338.5	338.5	338.5	338.5
6	399.9	399.8	399.6	399.6	399.6
7	455.2	455.0	455.1	455.1	455.1
8	503.8	503.7	503.6	503.6	503.6
9	545.9	546.5	546.1	546.1	546.0
10	582.3	584.6	584.5	584.5	584.5

The above equation is obtained using

$$\int_{-L/2}^{+L/2} \cos(Xz)\cos(K_{zn}z)\,dz = S(X), \tag{16a}$$

where

$$S(X) = \begin{cases} 0, & X = K_{zm},\ m \neq n, \\ L/2, & X = K_{zm},\ m = n, \\ \dfrac{2(-1)^{n+1}XL^2}{(4n^2\pi^2 - X^2L^2)}\sin\left(\dfrac{XL}{2}\right), & X \neq K_{zm},\ m = 1,2,3,\ldots. \end{cases} \tag{16b}$$

The boundary condition on T_{rz} on the curved surface is satisfied by using the orthogonal property of $\sin(K_{zm}z)$ in Eq. (5). Subsituting the specified function $\hat{T}_{rz}$ in the genral expression for T_{rz}, multiplying both sides by $\sin(K_{zn}z)$, and integrating over z yields

$$-\frac{L}{2}\sum_{s=1}^{3} B_{ns}[(c_{44}^E + e_{15}\gamma_{ns})K_{rns} + c_{44}^E\phi_{ns}K_{zn}]J_1(K_{rns}a) = \int_{-L/2}^{L/2} \hat{T}_{rz}\sin(K_{zn}z)\,dz, \quad n = 1,2,\ldots. \tag{17}$$

The boundary conditions on the potential are satisfied by using the orthogonal property of $J_0(k_{rn}r)$ in Eq. (7) of Ref. 18. Substituting the specified values of potential in Eq. (2), multiplying both side by $rJ_0(k_{rn}r)$, and integrating over r yields

$$A\left[-\frac{e_{33}a^2}{2\varepsilon_{33}^S}\sin\left(\frac{K_{z0}L}{2}\right)\right] + D\left[-\frac{La^2}{4}\right] + E\left(\frac{a^2}{2}\right) = 0 \tag{18a}$$

when $n=0$ for the bottom surface

$$A\left[\frac{e_{33}a^2}{2\varepsilon_{33}^S}\sin\left(\frac{K_{z0}L}{2}\right)\right] + D\left[\frac{La^2}{4}\right] + E\left(\frac{a^2}{2}\right) = \int_0^a \Phi_0 r\,dr = \Phi_0\left(\frac{a^2}{2}\right), \tag{18b}$$

when $n=0$ for the top surface, and

$$\sum_{s=1}^{3} A_{ns}\chi_{ns}\sin\left(\frac{k_{zns}L}{2}\right)\frac{a^2}{2}J_0^2(k_{rn}a) = 0, \tag{18c}$$

when $n=m=1,2\ldots$, for both top and bottom flat surfaces because they are equipotential surfaces.

The boundary condition on D_r on the curved surface is satisfied by using the orthogonal property of $\sin(k_{zm}z)$ in Eq. (7). Substituting the specified function $\hat{D}_r$ in the general expression for D_r multiplying both sides by $\sin(k_{zn}z)$, and integrating over z yields

$$\frac{L}{2}\sum_{s=1}^{3} B_{ns}[(\varepsilon_{11}^s\gamma_{ns} - e_{15})K_{rns} - e_{15}\phi_{ns}K_{zn}]J_1(K_{rns}a) = \int_{-L/2}^{L/2} \hat{D}_r\sin(K_{zn}z)\,dz, \quad n = 1,2,\ldots. \tag{19}$$

The open-circuit boundary condition in Eq. (11) is satisfied by equating the expression for the current in Eq. (8) to zero, i.e.,

$$2Be_{31}J_1(K_{r0}a) - D\varepsilon_{33}^S a = 0. \tag{20}$$

TABLE III. Material properties of PZT4 with internal losses. The real parts of the complex coefficients are from Wilson.[26]

Material property	Value
ρ (kg/m^3)	7500
ε_{11}^T (pF/m)	$1475\times 8.854(1-j0.004)$
ε_{33}^T	$1300\times 8.854(1-j0.004)$
s_{11}^E (TPa)$^{-1}$	$12.3(1-j0.005)$
s_{12}^E	$-4.05(1-j0.005)$
s_{13}^E	$-5.31(1-j0.005)$
s_{33}^E	$15.5(1-j0.005)$
s_{44}^E	$39.0(1-j0.005)$
d_{31} (pC/N)	$-123(1-j0.005)$
d_{33}	$289(1-j0.005)$
d_{15}	$496(1-j0.005)$
ν^D (m/s)	$4533+j10.22$
ν^E	$4305+j10.76$

TABLE IV. Critical frequencies and local maxima and minima in conductance, G, and susceptance, B, of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 1. Results are obtained using $I=J=20$ in ATILA and $M_r=M_z=20$ in the present method. However, * indicates that $M_r=M_z=30$ is used in the present method.

	Conductance				Susceptance							
	f_s (kHz)		G_{max} (mS)		$f_{-1/2s}$ (kHz)		B_{max} (mS)		$f_{1/2s}$ (kHz)		B_{min} (mS)	
No.	ATILA	Present method	ATILA	Present method	ATILA	Present method	ATILA	Present method	ATILA	Present method	ATILA	Present method
1	140.01	140.01	7.081	7.082	139.66	139.66	3.577	3.577	140.35	140.35	−3.504	−3.505
2*	219.68	219.70	0.461	0.461	219.18	219.19	0.261	0.261	220.19	220.21	−0.199	−0.199
3	286.14	286.15	0.0431	0.0431	285.47	285.48	0.0840	0.0840	286.82	286.83	0.0420	0.0420
4	340.77	340.77	0.0145	0.0145	339.97	339.98	0.0934	0.0934	341.56	341.57	0.0800	0.0800
5	398.80	398.82	0.00443	0.00441	398.01	398.04	0.1119	0.1119	399.59	399.61	0.1086	0.1086

For each special case the relevant Eqs. are truncated, combined, and expressed in matrix from as

$$[\mathbf{F}]\{\mathbf{X}\}=\{\mathbf{G}\}, \tag{21a}$$

where

$$\{\mathbf{X}\}^T=[A,B,D,E,A_{11},A_{12},A_{13},A_{21},A_{22},A_{23} \ldots A_{M_r1},A_{M_r2},A_{M_r3},B_{11},B_{12},B_{13},B_{21},B_{22},B_{23} \ldots B_{M_z1},B_{M_z2},B_{M_z3}]. \tag{21b}$$

$[\mathbf{F}]$ is a square matrix of size $[4+3(M_r+M_z),4+3(M_r+M_z)]$, and M_r and M_z are now finite. The elements of the column matrix $\{\mathbf{G}\}$ are nonzero only when the corresponding boundary conditions, i.e., stress boundary conditions and electrical boundary conditions on the surfaces, are nonzero. Solving Eq. (21) yields the weights in Eq. (2) and all fields of interest are determined by using them. The number of equations is equal to the number of unknown weights irrespective of the values of M_r and M_z.

1. Case 1

The applied potential Φ_0 is 1, shear stress is zero on all surfaces, axial stress is zero on the flat surfaces, and radial stress and normal component of electric displacement are zero on the curved surface, i.e., $T_{rz}=0$ on $0\leqslant r\leqslant a$, $z=\pm L/2$,and $r=a$, $|z|\leqslant L/2$, $T_{zz}=0$ on $0\leqslant r\leqslant a$, $z=\pm L/2$, $T_{rr}=0$ on $r=a$, $|z|\leqslant L/2$, $D_r=0$ on $r=a$, $|z|\leqslant L/2$, $\Phi=0$ on $0\leqslant r\leqslant a$, $z=-L/2$, and $\Phi=1$ on $0\leqslant r\leqslant a$, $z=+L/2$.

Equations (14), (17), (12a), (12b), (15a), (15b), (19), and (18a)–(18c) are equivalent to the boundary conditions and are combined in the same sequence. It therefore follows that the only non-zero element of $\{\mathbf{G}\}$ is

$$G_{2M_r+3M_z+4}=\int_0^a \Phi_0 r\,dr=\frac{a^2}{2} \tag{22}$$

2. Case 2

The boundary conditions on T_{rz}, T_{zz}, D_r, and Φ are the same as those in case 1. However, the normal stress on the curved surface, T_{rr}, is nonzero only in a band, i.e.,

$$T_{rr}=\begin{cases} \Gamma \text{ on } r=a, & |z|\leqslant L/(2N), \\ 0 \text{ on } r=a, & L/(2N)<|z|\leqslant L/2. \end{cases} \tag{23}$$

Equations (14), (17), (12a), (12b), (15a), (15b), (19), and (18a)–(18c) are again combined in the same sequence. It therefore, follows that the non-zero elements of $\{\mathbf{G}\}$ are

TABLE V. Conductance, G, and susceptance, B, of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 1.

Frequency (kHz)	G (μS)					
	ATILA[20]		Present method			
	$I=J=20$	$I=J=40$	$M_r=M_z=10$	$M_r=M_z=20$	$M_r=M_z=50$	$M_r=M_z=100$
100	0.6227	0.6227	0.6227	0.6227	0.6227	0.6227
200	0.7033	0.7033	0.6988	0.7021	0.7031	0.7032
300	0.4662	0.4661	0.4670	0.4663	0.4661	0.4661
400	1.966	1.916	2.233	1.991	1.925	1.916
500	219.2	220.1	219.7	220.0	220.1	220.1
1000	121.7	136.2	136.1	136.2	136.2	136.2
	B (μS)					
100	83.00	83.00	83.00	83.00	83.00	83.00
200	25.85	25.86	25.71	25.82	25.85	25.85
300	67.57	67.57	67.56	67.57	67.57	67.57
400	108.8	108.8	108.8	108.8	108.8	108.8
500	171.7	166.9	168.7	167.1	166.6	166.6
1000	338.3	298.8	297.6	296.0	295.4	295.3

TABLE VI. Axial displacement, U, at $r=0$, $z=L/2$ of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 1.

Frequency (kHz)	Re(U) (pm)					
	ATILA[20]		Present method			
	$I=J=20$	$I=J=40$	$M_r=M_z=10$	$M_r=M_z=20$	$M_r=M_z=50$	$M_r=M_z=100$
100	−299.4	−299.4	−299.0	−299.3	−299.3	−299.4
200	250.4	250.4	248.0	249.6	250.2	250.3
300	60.85	60.83	60.81	60.78	60.81	60.82
400	225.8	224.3	230.8	226.5	224.6	224.3
500	96.77	78.12	86.47	79.31	77.20	76.88
1000	23.27	24.61	25.14	25.28	25.40	25.42
	Im(U) (pm)					
100	3.073	3.073	3.068	3.071	3.071	3.073
200	−2.388	−2.388	−2.319	−2.367	−2.384	−2.387
300	2.099	2.097	2.109	2.099	2.097	2.097
400	169.3	164.0	200.8	172.8	165.1	164.0
500	−848.5	−853.9	−864.2	−857.3	−854.9	−854.4
1000	−1.701	21.88	8.28	19.27	22.85	23.45

$$G_{2M_r+M_z+2}=\int_{-L/(2N)}^{L/(2N)}\hat{T}_{rr}\,dz=\frac{L\Gamma}{N}, \tag{24a}$$

$$G_{2M_r+M_z+2+n}=\int_{-L/(2N)}^{L/(2N)}\hat{T}_{rr}\cos(K_{zn}z)\,dz$$
$$=\frac{L\Gamma}{n\pi}\sin\left(\frac{n\pi}{N}\right),\quad n=1,2,...,M_z, \tag{24b}$$

and Eq. (22). It is noted that the real part of the input current can be negative at some frequencies because of the applied stress. This, however, cannot occur in case 1.

3. Case 3

The boundary conditions on T_{rz}, T_{rr}, and D_r are the same as those in case 1; unit axial stress is applied on the flat

TABLE VII. Axial displacement, U, at $r=0$, $z=$L/2 of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 2.

Excitation N	Frequency (kHz)	Re(U) (pm)					
		ATILA[20]		Present method			
		$I=J=20$	$I=J=40$	$M_r=M_z=10$	$M_r=M_z=20$	$M_r=M_z=50$	$M_r=M_z=100$
1	200	770.0	777.0	770.4	774.8	776.5	776.8
	500	273.5	218.2	246.5	222.8	215.7	214.6
	1000	−20.85	−47.89	−43.40	−46.89	−47.91	−48.08
2	200	271.1	271.2	273.9	272.1	271.4	271.2
	500	−328.0	−258.4	−279.7	−259.5	−254.3	−253.5
	1000	138.3	175.4	169.6	176.4	177.9	178.2
20	200	250[a]	169.7[b]	174.4	171.1	170.1	169.9
	500	92[a]	−1804[b]	−1939	−1829	−1794	−1789
	1000	22[a]	494.4[b]	464.1	493.3	509.6	512.2
		Im(U) (pm)					
1	200	−6.052	−6.054	−5.871	−5.997	−6.041	−6.050
	500	−2479	−2492	−2519	−2501	−2495	−2493
	1000	19.49	82.25	48.66	76.20	85.19	86.69
2	200	13.23	13.24	13.04	13.17	13.22	13.23
	500	3255	3280	3329	3296	3285	3283
	1000	−13.90	−6.282	−11.66	−7.611	−6.288	−6.070
20	200	−2.4[a]	18.63[b]	18.28	18.53	18.61	18.62
	500	−850[a]	13609[b]	13779	13662	13624	13617
	1000	3.6[a]	398.9[b]	137.3	342.7	409.1	420.2

[a]625 elements.
[b]2500 elements.

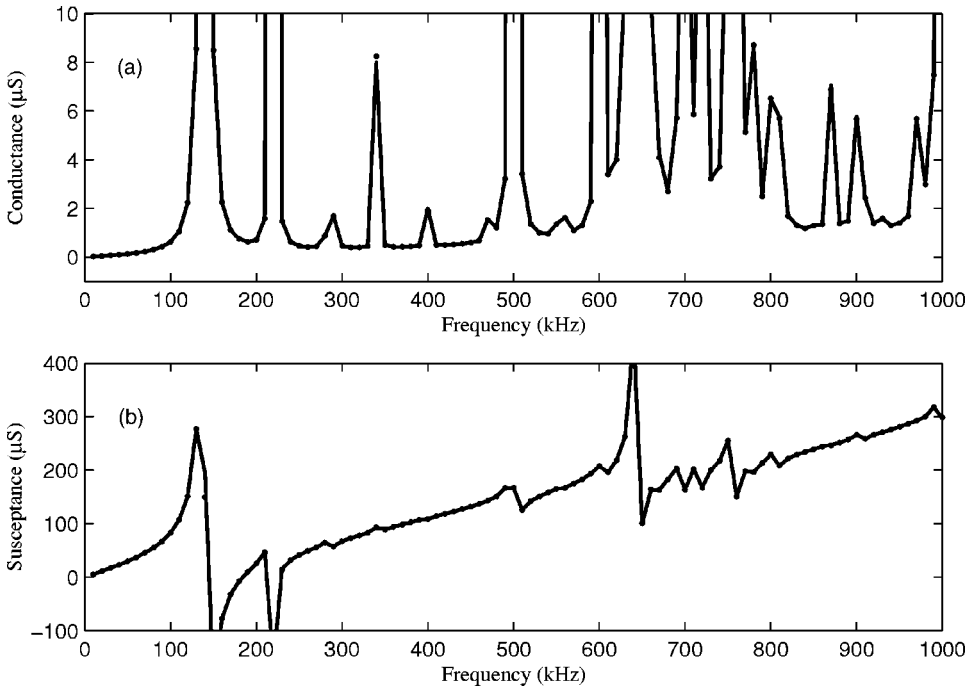


FIG. 2. (a) Conductance and (b) susceptance of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 1. Solid line: present method, dots: ATILA.

surfaces, i.e., $T_{zz}=1$ on $0\leqslant r\leqslant a$, $z=\pm L/2$, potential is zero on the bottom flat surface, and current is zero because of open-circuit condition.

Equations (14), (17), (12a), (12b), (15a), (15b), (19), (18a), (20), and (18c) are equivalent to the boundary conditions. They are combined in the same sequence. It, therefore, follows that the only nonzero element of $\{\mathbf{G}\}$ is

$$G_{M_r+M_z+1}=\int_0^a \overline{T}_{zz} r\, dr=\frac{a^2}{2}. \tag{25}$$

III. NUMERICAL RESULTS AND DISCUSSION

Numerical results obtained using the present method are compared with those obtained using ATILA (version 5.2.2) and other published results to illustrate the accuracy and rate of convergence of the present method. Values of certain critical frequencies, current, displacement, and potential are presented. In ATILA, second-order, axisymmetric, rectangular elements are used. The elements are equisized unless otherwise indicated. ATILA results are obtained using $I\times J$ elements, i.e., I elements in the axial direction and J elements in the radial direction. The number of finite elements that is necessary for accurate analysis depends on the frequency and spatial distribution of the excitation and is determined though convergence studies. Lines and dots are used to show all results obtained using the present approach and ATILA, respectively. All results shown here are for a cylinder of length 10 mm and radius 5 mm unless otherwise indicated.

The properties of PZT5A used to compute the resonance frequencies of a cylinder of length 2.03 mm and radius 20.05 mm ($L/a\approx 1/10$) and zero stress on all the surfaces are shown in Table I. The piezoelectric coefficients are all real and the input electrical conductance, G, of the cylinder is therefore zero. The input electrical susceptance, B, is discontinuous at the resonance frequencies. Guo and Cawley[25] used the finite element method and these dimensions to compute to four significant digits the resonance frequencies. They also listed several material properties of PZT5A. However, they did not state which set of ten coefficients needed to completely define a piezoelectric ceramic is used in the computations. Therefore, the ten coefficients taken from their list and used here are shown in Table I. The resonance frequencies computed by Guo and Cawley, those computed using 4×40 elements in ATILA, and $M_r/M_z=10$ in the present method are shown in Table II. The coefficients input to ATILA are computed using the coefficients shown in Table I

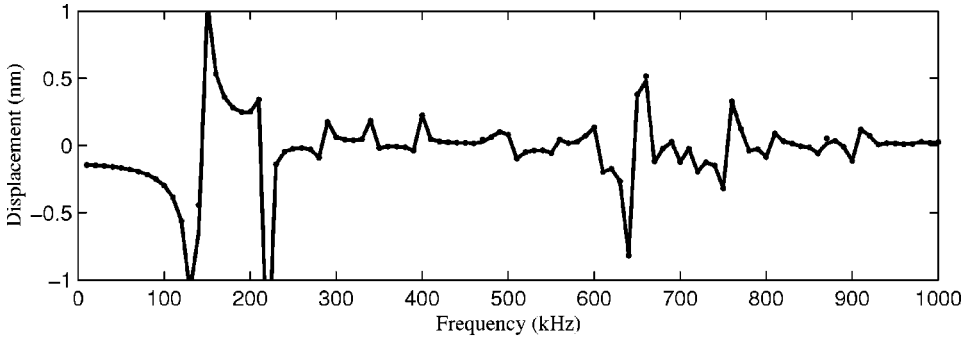


FIG. 3. Real part of axial displacement, U, at $r=0$, $z=L/2$ of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 1. Solid line: present method, dots: ATILA.

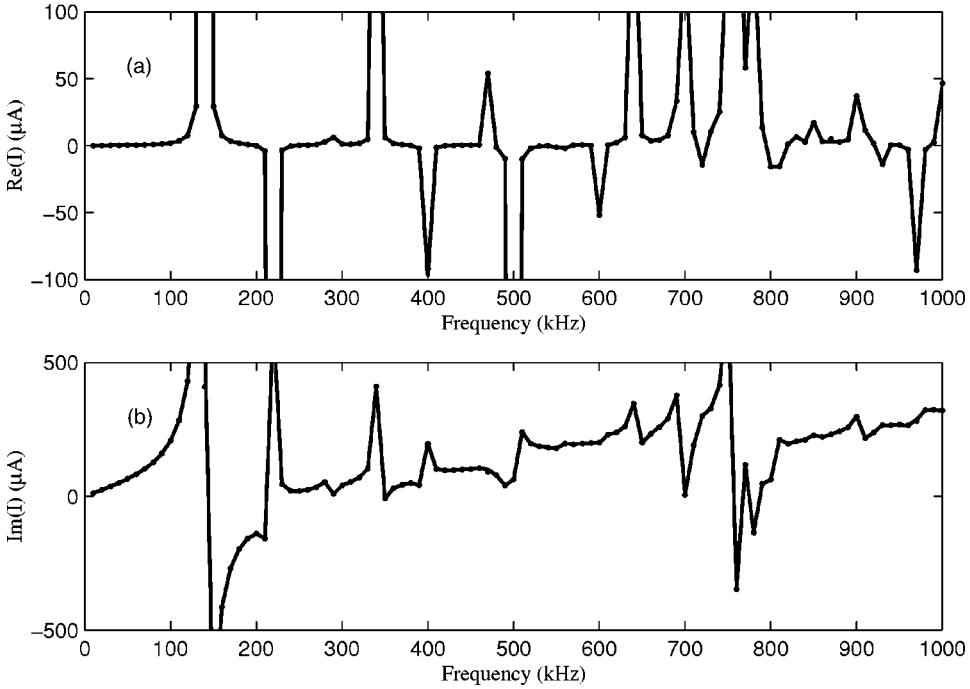


FIG. 4. Real part of current, I, of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 2, $N=2$. Solid line: present method, dots: ATILA.

and are not taken from Ref. 25. The agreement between the three methods is good. The maximum difference between the ten lowest resonance frequencies computed using ATILA and the present method is less than 0.05% even when $M_r=20$ and $M_z=2$.

The properties of a piezoelectric material with complex coefficients are shown in Table III. The real parts of these coefficients are those of PZT4[26] and the imaginary parts satisfy the conditions derived by Holland.[27] In all further illustrations, these properties and a cylinder of length 10 mm and radius 5 mm are used. All results presented in the figures are at 10, 20, 30,...,1000 kHz and are obtained using 40×40 elements in ATILA and $M_r=M_z=20$ in the present method. The effect of increasing the number of elements and M_r and M_z values has also been carried out and is shown in Tables V–VIII.

Results for case 1 are shown in Tables IV–VI, and Figs. 2 and 3. Critical frequencies of a cylinder with zero stress on all the surfaces are presented in Table IV. The frequency, f_s, at which G reaches a local maximum, the maximum value of G, $G_{\max}$; and the frequencies $f_{-1/2s}$ and $f_{1/2s}$ at which B reaches a local maximum and minimum, respectively; and the maximum and minimum values of B, $B_{\max}$, and $B_{\min}$, respectively computed using ATILA and the present method are shown. In ATILA, 40×40 elements are used. In the present method, $M_r=M_z=20$ is used except to compute the critical frequencies near 219 kHz for which $M_r=M_z=30$ is used. In the neighborhood of 219 kHz, the magnitude of the displacement is maximum at $r=a$, $z=L/2$—indicating that the response is largely due to excitation of an edge mode.[12] The agreement is excellent and the difference between the critical frequencies computed using the two methods is less than 0.01% for all the five lowest frequencies. Slower convergence rates have been reported for edge modes in other investigations also.[11,19] The term in braces in Eq. (8) is found to be zero within computational accuracy, thereby confirming that the current is independent of the axial coordinate.

The input electrical admittance of the cylinder, computed at 10, 20, 30,...,1000 kHz using ATILA and the present method is shown in Fig. 2 for case 1. The scale of the Y axis is chosen to highlight the agreement at nonresonance frequencies because the values at resonance are compared in Table IV. Values of G and B at a few frequencies are shown in Table V. They are shown for various values of M_r and M_z and it is seen that convergence is very rapid in the present method. Results obtained using 20×20 and 40×40 elements in ATILA are also shown in Table V. In the former case, the element size is 0.5×0.25 mm^2 and the longer side is about

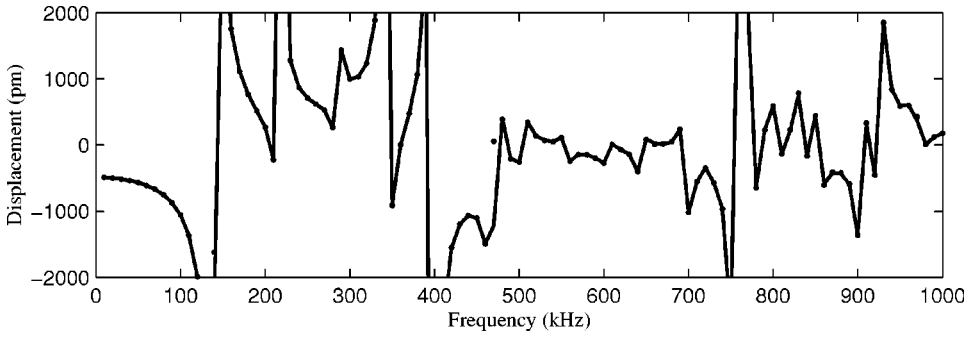


FIG. 5. Real part of axial displacement, U, at $r=0$, $z=L/2$ of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 2, $N=2$. Solid line: present method, dots: ATILA.

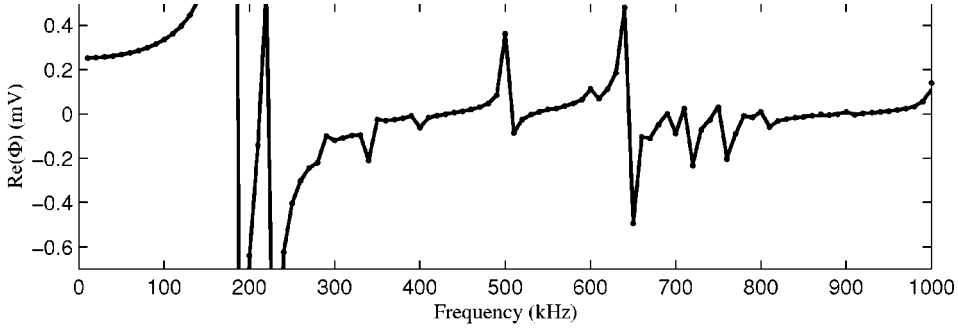


FIG. 6. Real part of potential, Φ, at $z=L/2$ of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 3. Solid line: present method, dots: ATILA.

one-tenth of the wavelengths $2\pi/K_{z0}$ and $2\pi/k_{r0}$ at 1 MHz. However, doubling the number of elements causes a significant change in G and B at 1 MHz and there is better agreement with the present method.

The axial displacement at $r=0$, $z=L/2$ are shown in Fig. 3 and Table VI for case 1. In the present method, rate of convergence is good for displacement but is better for the admittance. This is due to two reasons. (1) Admittance is computed by evaluating an integral over the surface area of the electrodes and thus represents an average effect. As noted earlier, admittance is dependent only on the coefficients B and D and is independent of the coefficients A_{ms} and B_{ms}. (2) Even though U is expressed in terms of a point-wise complete set of functions, $J_0(k_{rm}r)$, $m=0,1,2,...$, $\partial J_0(k_{rm}r)/\partial\, dr$ are all zero at $r=0$. Therefore, the rate of convergence at $r=0$ is slow as observed earlier.[18] In ATILA, doubling the number of elements has a significant effect at 1 MHz especially on the imaginary part of displacement. At low frequencies, U on the flat surfaces is approximately equal to the static displacement $-d_{33}/2=-1.445\times10^{-10}+j7.225\times10^{-13}$ m. The static result is obtained by using $T_{rr}=T_{\theta\theta}=T_{zz}=T_{rz}=0$ everywhere in the cylinder.

Results for case 2 ($N=1$ and $\Gamma=5\times10^3$; $N=2$ and $\Gamma=10^4$; and $N=20$ and $\Gamma=10^5$) are shown in Figs. 4 and 5 and in Table VII. In this case, a potential difference of 1 V is applied across the electrodes and the curved surface is simultaneously stressed. Input current is shown in Fig. 4 for $N=2$ and $\Gamma=10^4$, i.e., radial stress is applied over the central half of the cylinder, at 10, 20, 30,...,1000 kHz. The value of Γ is chosen such that the effect of the applied potential and the applied stress are comparable at low frequencies. In contrast to Fig. 2, the real part of the current is negative at some frequencies—a consequence of the application of mechanical excitation. The real and imaginary parts of the current computed using ATILA and the present method are in good agreement at all frequencies. Axial displacement is shown in Fig. 5 for $N=2$ and in Table VII for $N=1$, 2, and 20. It is seen from Table VII that more terms are required for convergence as N increases. For $N=1$, at low frequencies, U on the flat surfaces is approximately equal to the static displacement $(2s_{13}^E\Gamma-d_{33}\Phi_0/L)L/2\approx-4.10\times10^{-10}+j2.05\times10^{-12}$ m. The static result is obtained by using $T_{rr}=T_{\theta\theta}=\Gamma$ and $T_{zz}=T_{rz}=0$ everywhere in the cylinder. For $N=20$, instead of the 20×20 and 40×40 elements used in the other cases, ATILA results are obtained using a total of 625 and 2500 elements, respectively. In the latter mesh, a fine mesh of 900 elements is used in the regions $r\leqslant a/10$ and $|z|\leqslant L/40$, and 1600 elements are used in the remaining regions. Good agreement between ATILA and the present method is obtained even when the load is localized and the frequency is high when 2500 elements are used. However, when only 625 elements are used, only the results at 200 kHz are approximately correct.

Results for case 3 are shown in Fig. 6 and Table VIII. The potential difference generated by application of uniform unit stress on the flat surfaces is shown. The real and imaginary components of potential computed using ATILA and the present method are in good agreement. The real and imaginary parts of the potential have converged in the present method even at 1 MHz. Better agreement can be obtained at such frequencies by using more elements in ATILA.

TABLE VIII. Potential, Φ, at $z=$L/2 of a PZT4 cylinder of radius 5 mm and length 10 mm with internal losses for case 3.

Frequency (kHz)	Re(Φ) (mV)					
	ATILA[20]		Present method			
	$I=J=20$	$I=J=40$	$M_r=M_z=10$	$M_r=M_z=20$	$M_r=M_z=50$	$M_r=M_z=100$
200	−0.6396	−0.6396	−0.6460	−0.6413	−0.6398	−0.6396
500	0.3609	0.3631	0.3624	0.3630	0.3632	0.3632
1000	0.1388	0.1125	0.1112	0.1096	0.1090	0.1090
	Im(Φ) (mV)					
200	−0.0281	−0.0281	−0.0282	−0.0281	−0.0281	−0.0281
500	−0.2149	−0.2194	−0.2179	−0.2192	−0.2196	−0.2197
1000	−0.1533	−0.2037	−0.2049	−0.2067	−0.2074	−0.2075

Only the real parts of functions are compared in Figs. 3, 5, and 6. Similar good agreement is obtained for the imaginary parts also.

IV. CONCLUSIONS

An accurate method is presented to determine the forced responses of axially polarized, finite cylinders with internal losses. In this method, exact solutions to the governing equations are used and all field variables (displacements, potential, stress, and electric displacement) are expressed in terms of complete sets of functions. Therefore, arbitrary boundary conditions can be satisfied.

The method is illustrated by presenting the resonance frequencies for free vibration and the responses to a variety of excitations. Numerical values of the complex input electrical admittance and the complex displacement are computed and compared with finite element results. For cylinders with internal losses, the frequency, f_s, at which the conductance, G reaches a local maximum, the maximum value of G, $G_{\max}$, and the frequencies $f_{-1/2s}$ and $f_{1/2s}$ at which the susceptance B reaches a local maximum and minimum, respectively, and the maximum and minimum values of B, $B_{\max}$, and $B_{\min}$, respectively, are all in excellent agreement with ATILA. For a cylinder of length 10 mm and radius 5 mm, the five lowest of each of the above critical frequencies computed using ATILA and the present method differ by less than 0.01%. These frequencies are of interest in the design of devices and in the characterization[28] of piezoelectric material. The axial displacement at the center of the flat surfaces is also in good agreement with ATILA even when the excitation is localized and the frequency is high. The comparison is made at this point because convergence, in the present method, is relatively slow at $r=0$. For most practical applications, less than ten terms in the present method will be sufficient.

The method can be extended to analyze hollow cylinders by using appropriate complete sets of functions. It can also be extended to analyze Langevin transducers, piezocomposites, and other devices.

ACKNOWLEDGMENT

The authors thank the Director, Naval Physical and Oceanographic Laboratory, for granting permission to publish this paper and the Production Editor, Ms. Barbara Graham for helping us to reduce its length.

[1] H. Chung, "Free vibration analysis of circular cylindrical shells," J. Sound Vib. **74**(3), 331–350 (1981).
[2] J. T.-S. Wang and C.-C. Lin, "A method for exact series solutions in structural mechanics," J. Appl. Mech. **66**, 380–387 (1999).
[3] S. Hurlebaus, L. Gaul, and J. T.-S. Wang, "An exact series solution for calculating the eigenfrequencies of orthotropic plates with completely free boundary," J. Sound Vib. **244**(5), 747–759 (2001).
[4] P. Bisegna and F. Maceri, "An exact three-dimensional solution for simply supported rectangular piezoelectric plates," J. Appl. Mech. **63**, 628–638 (1996).
[5] S. Kapuria, P. C. Dumir, and S. Sengupta, "An exact axisymmetric solution for a simply supported piezoelectric cylindrical shell," Arch. Appl. Mech. **67**, 260–273 (1997).
[6] H. J. Ding, R. Q. Xu, and W. Q. Chen, "Free vibration of transversely isotropic piezoelectric circular cylindrical panels," Int. J. Mech. Sci. **44**, 191–206 (2002).
[7] J. N. Sharma, "Three-dimensional vibration analysis of a homogenous transversely isotropic thermoelastic cylindrical panel," J. Acoust. Soc. Am. **110**, 254–259 (2001).
[8] R. Holland and E. P. EerNisse, *Design of Resonant Piezoelectric Devices*, Research Monograph No. 56 (MIT, Cambridge, 1969); "Variational evaluation of admittances of multielectroded three dimensional piezoelectric structures," IEEE Trans. Sonics Ultrason. **SU-15**, 119–132 (1968).
[9] J. R. Hutchinson, "Axisymmetric vibrations of a free finite length rod," J. Acoust. Soc. Am. **51**, 233–240 (1972).
[10] M. C. Junger, "The sound field in a finite cylindrical shell," J. Acoust. Soc. Am. **77**, 1610–1612 (1985).
[11] R. Huang, P. C. Y. Lee, W. S. Lin, and J. D. Yu, "Extensional, thickness-stretch, and symmetric thickness-shear vibrations of piezoceramic disks," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**(11), 1507–1515 (2002).
[12] E. A. G. Shaw, "On the resonant vibrations of thick barium titanate disks," J. Acoust. Soc. Am. **28**, 38–50 (1956).
[13] E. P. EerNisse, "Variational method for electroelastic vibration analysis," IEEE Trans. Sonics Ultrason. **SU-14**(4), 153–160 (1967).
[14] D. D. Ebenezer and R. Ramesh, "Exact analysis of axially polarized piezoelectric ceramic cylinders with certain uniform boundary conditions," Curr. Sci. **85**(8), 1173–1179 (2003).
[15] C. Prasad and R. K. Jain, "Vibrations of transversely isotropic cylindrical shells of finite length," J. Acoust. Soc. Am. **38**, 1006–1009 (1965).
[16] H. S. Paul, "Vibrations of circular cylindrical shells of piezoelectric silver iodide crystals," J. Acoust. Soc. Am. **40**, 1077–1080 (1966).
[17] R. Ramesh and D. D. Ebenezer, "Analysis of axially polarized piezoceramic rings," Fourth Asian Meeting on Ferroelectrics, 12–15 December, Bangalore, India (2003), p. 38.
[18] D. D. Ebenezer and R. Ramesh, "Analysis of axially polarized piezoelectric cylinders with arbitrary boundary conditions on the flat surfaces," J. Acoust. Soc. Am. **113**, 1900–1908 (2003).
[19] D. D. Ebenezer, K. Ravichandran, and C. Padmanabhan, "Forced vibrations of solid elastic cylinders," J. Sound Vib. **282(3–5)**, 991–1007 (2005).
[20] J. N. Decarpigny, J. C. Debus, B. Tocquet, and D. Boucher, "In-air analysis of piezoelectric tonpilz transducers in a wide frequency band using a mixed finite element-plane wave method," J. Acoust. Soc. Am. **78**, 1499–1507 (1985). ATILA is developed by Acoustics Laboratory, I.S.E. N., Lille Cedex, France.
[21] A. Iula, R. Carotenuto, and M. Pappalardo, "An approximated 3-D model of the Langevin transducer and its experimental validation," J. Acoust. Soc. Am. **111**, 2675–2680 (2002).
[22] Y.-C. Lin and C.-C. Ma, "Experimental measurement and numerical analysis on resonance characteristics of piezoelectric disks with partial electrode designs," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 937–947 (2004).
[23] D. A. Berlincourt, D. R. Curran, and H. Jaffe, "Piezoelectric and piezomagnetic mateials and their functions in transducers," in *Physical Acoustics Vol. IA*, edited by W. P. Mason (Academic, New York, 1964).
[24] M. Abramowitz and I. A. Stegun (Eds.), *Handbook of Mathematical Functions* (Dover, New York, 1965), Eqs. 11.3.20, 11.3.29, and 11.4.5.
[25] N. Guo and P. Cawley, "Measurement and prediction of the frequency spectrum of piezoelectric disks by modal analysis," J. Acoust. Soc. Am. **92**(6), 3379–3388 (1992).
[26] O. B. Wilson, *Introduction to the Theory and Design of Sonar Transducers* (Peninsula, Los Altos, CA, 1988).
[27] R. Holland, "Measurement of piezoelectric phase angles in a ferroelectric ceramic," IEEE Trans. Sonics Ultrason. **SU-17**(2), 123–124 (1970).
[28] D. D. Ebenezer, "Determination of complex coefficients of radially polarized piezoelectric ceramic cylindrical shells using thin shell theory," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**(10), 1209–1215 (2004).

On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach[a)]

Sang-Myeong Kim[b)] and Wonjae Choi[c)]
Department of Mechatronics, Gwangju Institute of Science and Technology, 1 Oryong-dong, Puk-gu, Gwangju, 500-712, Korea

(Received 31 May 2004; revised 4 April 2005; accepted 4 April 2005)

In-head localization of sound images is a critical problem in headphone reproduction. The paper investigates the degree of externalization in terms of the distance of auditory images for various synthesis and reproduction cases. An effective binaural headphone system was constructed by way of binaural synthesis using head-related impulse responses and individual headphone equalization using Wiener filter theory. The headphone system designed had an average reproduction performance error of 2.4% for five subjects with a random noise input, and was used to perform some subjective tests with a set of virtual sources equally spaced and distanced from the center of each subject's head in the horizontal plane. The effects of individual and nonindividual binaural syntheses and those of equalized and nonequalized reproductions were separately investigated. In the tests, each subject was instructed to indicate the distance of auditory images. The results obtained demonstrate that individual equalization is important for externalization, and individual synthesis is important for consistent distance perception. Thus, a combined use of both individual equalization and individual synthesis resulted in externalized sound images of a consistent distance.
 [DOI: 10.1121/1.1921548]

PACS numbers: 43.38.Vk, 43.38.Md, 43.66.Pn, 43.66.Qp, 43.66.Yw [AK] Pages: 3657–3665

I. INTRODUCTION

The ultimate goal of binaural technology is to give the listener a vivid and realistic impression, as if he or she were in the actual place of the acoustic events. Two steps are generally required for this to occur. First, binaural signals containing all the spatial information of acoustic events with respect to a ''target listener'' must be prepared from either recording or synthesizing. Second, an effective binaural (or authentic) reproduction system should be designed to reproduce the signals authentically at the eardrums of a ''destination listener,'' who is in a different environment equipped with loudspeakers or headphones. Headphone reproduction is often preferred due to its powerful characteristics regarding perfect channel separation and head-oriented installation. It is thus free from those obstacles generally occurring in loudspeaker reproduction, such as cross-talk, sweet spot, and listening room problems. However, it also has some critical acoustical drawbacks apart from a possible feeling of discomfort that comes from wearing headphones. As headphones are worn on the listener's head, the auditory events tend to come inside the head (in-head localization) and front–back orientations are often reversed (front–back confusion). These can seriously harm the listening comfort and the clarity of a spatial auditory display.

In-head localization (IHL), as well as front–back confusion, is an inherent problem of headphones and has been known for more than half a century among psychoacoustic researchers. As reviewed by Blauert (1983, Secs. 2.1–3), there have been many hypotheses regarding the reasons: (1) differences in the sound transmission channels between real-life listening and headphone listening; (2) unnatural resonances of the headphone cavity; (3) unnatural loadings of the ear drum; (4) invariability of the headphone signals when the head is moved; (5) bone-transmitted sounds, etc. The first three hypotheses are closely related to each other and have gained more support. Although less general, IHL can also happen in loudspeaker reproduction, particularly when the listener's head is immobilized and no visual information of the source is presented (Blauert, 1983). Toole (1969) reported that a source in the median plane, such as that in the right front or rear, occasionally caused IHL for some listeners. The front source was more frequently judged to be inside than the rear source. IHL could also happen if two sources located at both sides of the ears emitted the same signal.

Whatever the reasons are for IHL in headphone reproduction, there have been many endeavors to resolve this. Sakamoto *et al.* (1976) tried to present somewhat externalized auditory illusions by adding enough artificial reverberation. A more proper treatment was pioneered by Wightman and Kistler (1989a, b) based on tackling the first hypothesis described above: differences in the sound transmission channels between real-life listening and headphone listening. In order to deliver the same realistic signals heard in the real world to a headphone listener, real-world signals were synthesized using head-related-transfer-functions (HRTFs) and presented using equalized headphones. The delivery was moderately successful in terms of both objective accuracy and subjective directional discriminations. However, the distance of sound images accounting for the degree of external-

[a)]Portions of this work were presented in ''Objective and subjective evaluations of 3D sound reproduction in a headphone,'' Inter-Noise 2003, Jeju, Korea, August 2003.

[b)]Present address: EMPA, Ueberlandstr. 129, Duebendorf, Switzerland; electronic mail: smkim123@hanmail.net
[c)]Present address: Digital Appliance Research Lab, LG Electronics, Inc., 327-23 Gasan-dong, Geumcheon-gu, Seoul, 153-802, Korea.

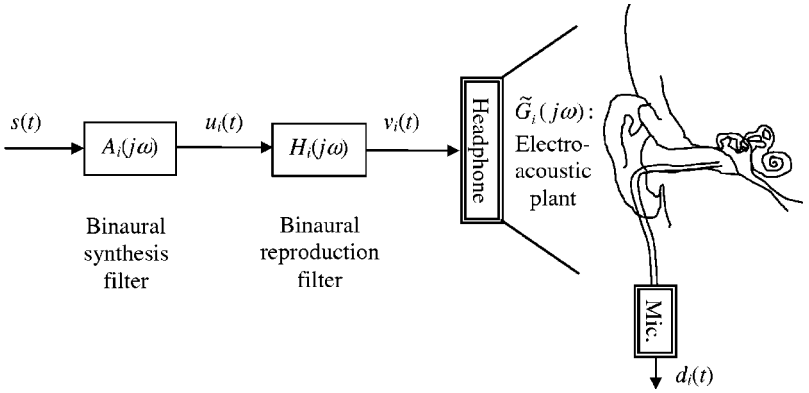


FIG. 1. Binaural technology in headphone reproduction: an original monaural sound $s(t)$, the binaurally synthesized sound $u_i(t)$, the headphone input signal $v_i(t)$, the probe microphone output signal $d_i(t)$, where the subscript $i=1,2$ denotes the left- or right-ear channel of a listener, respectively.

ization was not investigated and the equalizer employed was a rather simple one constructed by an inverse Fourier transform of the ratio between the acoustic responses recorded in the free field and under headphones. Later, Møller *et al.* (1996) took up a similar task, using binaural signals directly recorded from a subject's own head (individual recording) and other heads (nonindividual recordings). Headphones were individually equalized for each subject, and the method was a minimum phase IIR-filter approximation of the target amplitude response only, disregarding its phase response. Their subjective tests demonstrated that binaural headphone listening was as good as real-life listening in terms of distinguishing a number of ''predefined'' locations of sound sources. The results using nonindividual recordings were slightly worse for the sources in the median plane. Again, the degree of externalization was not considered. Hartmann and Wittenberg (1996) introduced a simple synthesis and equalization technique suitable for their specific task studying psycho-acoustical parameters of externalization of some simple sounds.

The success of the research activities in the above-described category is primarily determined by a proper use of headphone equalizers. The work presented here employs an optimal equalization method based on the Wiener filter approach and closely follows the systematic framework recently presented by Kim and Wang (2003) for the authentic reproduction of stereo sound in loudspeaker reproduction. The designed equalization filters are optimal in terms of reproducing the original signals from the sense of minimizing the mean square errors. With this framework, the reproduction performance errors can be easily quantified by a simple comparison between the original and reproduced signals in the time domain. However, it should be noted that such optimal filters do not necessarily bring optimal performances from the human perception point of view. Nonetheless, the Wiener filter method employed in this paper should be a good candidate on which one may rely, particularly when no golden method is known. After accuracy is ensured objectively to some degree, the test system is then used to conduct some subjective tests that are focused on the degree of externalization in terms of the distance of auditory images.

In the tests, the subjects were requested to identify the location of ''auditory events or images'' (where one feels), rather than to estimate that of ''actual sources'' (where one thinks). Such a distinction is consistent with that used by Blauert (1983, p. 46). One may interpret that the former has no fixed correct answer and is wholly conducted in the subject's internal auditory space. On the other hand, the latter generally has a correct answer from an objective sense and involves some previous experience and knowledge since it requires mapping the internal auditory space into the external acoustic space. The effects of individual and nonindividual binaural syntheses and those of equalized and nonequalized reproductions are investigated. The externalization performance obtained from the subjective tests was encouraging, which could support the validity of the test system used.

II. WIENER FILTER APPROACH

A. Binaural technology in headphone reproduction

Binaural technology in virtual sound is comprised of binaural synthesis and binaural reproduction. The former is needed to render spatial effects on an original sound signal by using synthesis filters such as HRTFs and room-transfer-functions. The latter is needed to authentically reproduce the synthesized signals to the listener's eardrums by using equalization (or inverse) filters that can cancel out the effect of a particular electro-acoustic reproduction path.

A complete implementation procedure for headphone reproduction is illustrated in block diagram form in Fig. 1, where binaural synthesis and reproduction techniques are implemented by way of a HRTF $A_i(j\omega)$ and an equalization filter $H_i(j\omega)$, respectively. The subscript $i=1,2$ denotes the left- or right-ear channel of a listener, respectively. The signal $s(t)$ is the original sound signal, $u_i(t)$ is the binaurally synthesized signal, $v_i(t)$ is the headphone input signal, and $d_i(t)$ is the probe microphone output signal. The real physical electro-acoustic system $\tilde{G}_i(j\omega)$ represents the acoustical eardrum response $d_i(t)$ to the electrical headphone input $v_i(t)$. The tilde in $\tilde{G}_i(j\omega)$ has been used to stress a real physical system that usually contains measurement noise and is often slightly nonlinear, and those without it represent ideal linear systems. Since there is a perfect channel separation between the left- and right-ear channels, only a single channel is illustrated in Fig. 1 for simplicity. One can draw the other channel easily, but note the same monaural signal $s(t)$ is also fed to that channel. To design these binaural synthesis and reproduction filters, some knowledge on sys-

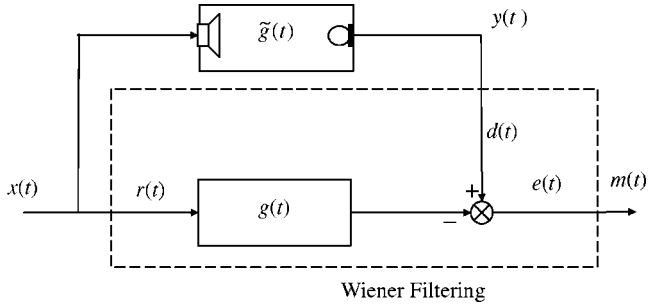


FIG. 2. Wiener's problem for finding an optimal linear filter $g(t)$ that minimizes the error signal $e(t)$ in the mean square sense when the received $r(t)$ and desired $d(t)$ signals are ergodic and stationary random. The real physical system containing measurement noise is denoted as $\tilde{g}(t)$ for distinction.

tem identification techniques is a prerequisite, which is summarized in the following section.

B. System modeling filter

Let a real physical system $\tilde{g}(t)$ be excited by a test signal $x(t)$ and its response be $y(t)$. Using both input and output signals, one can estimate a linearized model $g(t)$ of the system. This modeling process is depicted in Fig. 2, where the input–output relationship can be written as

$$y(t)=g(t)*x(t)+m(t), \tag{1}$$

where the asterisk $(*)$ denotes the linear convolution operator and $m(t)$ is the *modeling error*.

To obtain the optimal model for $\tilde{g}(t)$ that minimizes the mean square of the error $m(t)$, the Wiener filter can be used. There have been various system identification methods, but most fall in the category of Wiener filter (Haykin, 1996; Papoulis, 1984). The fundamental structure of the Wiener filter is indicated inside the dashed box in Fig. 2 with its typical notations [i.e., the received $r(t)$, desired $d(t)$, error $e(t)$ signals]. Their relations with the physical signals are: $r(t)=x(t)$, $d(t)=y(t)$, and $e(t)=m(t)$. The Wiener filter is the physically realizable optimal filter that minimizes the error $e(t)$ in the mean square sense when the desired $d(t)$ and received $r(t)$ signals are ergodic and stationary random.

When implemented by a finite impulse response (FIR) filter in the discrete time domain, the Wiener filter $g[k]$ is the solution of the discrete form of the Wiener–Hopf equation given by

$$\sum_{k=0}^{\infty} R_{rr}[n-k]g[k]=R_{rd}[n], \qquad n\geqslant 0, \tag{2}$$

where $R_{rr}[n]$ and $R_{rd}[n]$ are, respectively, the autocorrelation of the received signal $r[k]$ and the cross-correlation between the received signal and the desired signal $d[k]$. If the filter has I coefficients, then Eq. (2) can be expressed in matrix form as

$$\mathbf{Ag}=\mathbf{b}, \tag{3}$$

where the $(I\times I)$ autocorrelation matrix $\mathbf{A}$ is a Toeplitz matrix, $\mathbf{g}$ is the I length Wiener filter coefficient vector, and $\mathbf{b}$ is the I length cross-correlation vector (Haykin, 1996; Papoulis, 1984). Thus, the Wiener filter is given by

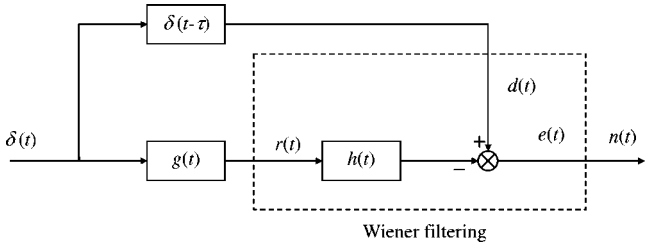


FIG. 3. Deconvolution of $g(t)$ using the inverse filter $h(t)$. The dashed box indicates a Wiener filter problem.

$$\mathbf{g}=\mathbf{A}^{-1}\mathbf{b}. \tag{4}$$

If the received signal is spectrally rich as in random processes, the matrix $\mathbf{A}$ is guaranteed to be positive definite. If not, one may add robustness, for example, by using $\mathbf{A}+\beta\mathbf{I}$ instead of $\mathbf{A}$, where β is a robustness coefficient and $\mathbf{I}$ is an identity matrix (Elliott, 2001).

There are two computationally intensive processes in Wiener filtering: one for the averaging process for estimating the correlation functions and the other for the inverse matrix calculation. When some special input signals are used, however, these burdens can be greatly relieved. First, when the input signal is a periodic sequence of a long maximum-length-sequence (MLS) (Vanderkooy, 1994), the autocorrelation matrix becomes an identity matrix so that no inverse process is required (i.e., $\mathbf{g}=\mathbf{b}$). Due to this character, MLSs are particularly welcome for acoustic systems that often have long reverberation times, as also used for the HRIR (head-related-impulse-response, which is the time domain representation of a HRTF) measurement described in the Appendix of this paper. Golay codes are another example in this category (Zahorik, 2000). Second, when the test signal and the system output signal are deterministic transient signals having no additive noise, no averaging process is required (Kim and Wang, 2003). This situation is called deterministic Wiener filtering while a case having random signals is known as random Wiener filtering.

C. Equalization filter

The deterministic Wiener filter can be effectively used for designing the equalization inverse filter $h(t)$ whose frequency response function corresponds to $H_i(j\omega)$ in Fig. 1. Once the optimal system model $g(t)$ is identified by the procedure in the previous section, this model is then used for designing its inverse filter $h(t)$. As illustrated in Fig. 3, this design task is a typical deconvolution problem that is to perfectly reproduce the original input signal at the output by deconvolving $g(t)$ using its inverse filter $h(t)$. In the diagram, a time delay in $\delta(t-\tau)$ is introduced in order to yield a causal inverse filter $h(t)$. An appropriate time delay τ should be endorsed for a satisfactory equalization due to the nonminimum phase characteristics of the system. It is again a Wiener filtering problem, as indicated by a dashed box in Fig. 3. The input–output relationship can be written as

$$\delta(t-\tau)=g(t)*h(t)*\delta(t)+n(t), \tag{5}$$

where $n(t)$ can be defined as the *inverse error*. The optimal equalization filter $h(t)$ can be obtained by the deterministic

Wiener filtering with setting $r(t)=g(t)$ and $d(t)=\delta(t-\tau)$. Since it inevitably involves a matrix inversion as given by Eq. (4), this Wiener filter equalization method is effective particularly when the system $g(t)$ and the equalizer $h(t)$ are realized with short FIR filters. If any of them is very long, an approximate but faster method may have to be considered, such as a robust frequency domain inverse method (Kirkeby *et al.*, 1998) or minimum phase approximations of the target amplitude response (Gardner and Martin, 1994; Møller *et al.*, 1996; Oppenheim and Schafer, 1989).

When the equalization filter is implemented as shown in Fig. 1, the final *reproduction error* can be calculated by comparing the binaurally synthesized signal $u_i(t)$ and the microphone measured signal $d_i(t)$ as given by

$$e_i(t)=d_i(t)-u_i(t-\tau), \tag{6}$$

where $i=1,2$. The modeling, inverse, and reproduction errors can be more simply quantified in terms of variances of signals, as is generally used in the Wiener filter theory. Thus, the corresponding performance errors are given in the Wiener filter framework by

$$J(\%)=\frac{E[e^2(t)]}{E[d^2(t)]}\times 100, \tag{7}$$

where $E[\circ^2]$ denotes the variance, $e(t)=m(t)$ and $d(t)=y(t)$ in system modeling, $e(t)=n(t)$ and $d(t)=\delta(t-\tau)$ in equalization, and $e(t)=e_i(t)$ and $d(t)=u_i(t-\tau)$ in reproduction. The notations $r(t)$, $d(t)$, and $e(t)$ have been used only for Wiener filtering and inside the dashed boxes in Figs. 2 and 3. They should not be confused with the physical variables in Fig. 1, for example, $d_i(t)$. It is important to notice that all the errors involved, such as modeling, inverse, and unpredictable repeatability errors are finally manifested in the form of reproduction error, and thus the reproduction performance error can be used as an objective accuracy measure of the binaural reproduction system. For the details of equalization filter design and error analysis, readers are referred to the work by Kim and Wang (2003).

III. HEADPHONE EQUALIZATION

A. Subjects and test conditions

Both objective and subjective tests were carried out with five male subjects between 27 and 37 years of age with normal hearing, whose initials were S, M, G, W and D, where S and W were the authors of this paper. All tests were conducted in a noise-proof recording studio booth (Sound & Noise 1818 model having 1.8 m×1.8 m×2.0 m internal volume) using an open-type headphone (Sennheiser HD600) and a probe microphone (Etymothic Research ER-7) inserted about 20 mm into the ear canal, as illustrated in Fig. 1. The test system was computer-controlled using the MATLAB data acquisition and signal processing toolboxes and a soundcard (Lynx II) as a front-end device for data acquisition and transmission at a sampling frequency of 44.1 kHz.

B. Headphone equalization

For the system modeling, the test signal was a random signal of length 44 100×2 samples. Only the right ear was tested and its corresponding equalization filter was also used for equalizing the other channel. This arrangement is arguable because the electro-acoustic response of one channel of the headphones will never be exactly the same as that of the other channel. Each channel response is dependent on various factors including the shape of the ear, the headphone fitting status on the ear and jaw, the location of the probe microphone inside the ear cannel, etc. However, some degree of symmetry was assumed mainly for the convenience of testing. This assumption was also convenient in this study because there was no guarantee for a consistent headphone fitting status during the whole course of tests including the successive subjective tests where the probe microphone was pulled out to avoid any discomfort while testing and indicating answers using a pencil, as will be described later. Since system responses were quite different across subjects, however, modeling and equalization were performed for each subject. The Wiener filter model for each system was obtained using the random Wiener filtering described in Eqs. (2)–(4) with the robustness coefficient $\beta=0$. Its corresponding equalization filter was designed using the deterministic Wiener filtering described in the previous section with a robustness coefficient of $\beta=10^{-4}$ to account for the effects of slight nonsymmetry between two channels and possible headphone fitting changes while testing. For both identification and equalization, a FIR filter length of 1024 was used and a 256-sample delay was given to τ in the inverse filter design.

Figures 4(a) and 4(b) show the modeled system $g_1(t)$ of subject W and its equalization filter response $h_1(t)$, respectively. Figure 4(c) shows the frequency domain representations of the modeled system $G_1(j\omega)$ (solid line) and the equalization filter $H_1(j\omega)$ (dashed). In addition, the complete channel response (dotted) computed by $G_1(j\omega)\cdot H_1(j\omega)$ shows a nearly flat response, which demonstrates an excellent equalization (in the time domain, it is an impulse-like signal).

In order to assess the accuracy of the reproduction system designed, the reproduction error was also examined for each subject. A description of the test procedure follows. Since $u_1(t)=s(t)$ in Fig. 1, the test random signal $s(t)$ was first convolved with the equalizer $h_1(t)$ and the resulting convolved random signal $v_1(t)$ was then transmitted to the headphone. By comparing the measured response $d_1(t)$ and the original signal $s(t)$, the reproduction error could be calculated using Eqs. (6), (7) with the setting $u_1(t-\tau)=s(t-\tau)$. The time delay τ could be numerically identified from the cross-correlation function between $s(t)$ and $d_1(t)$. Table I shows the modeling, inverse, and reproduction performance errors for all subjects. Less than 0.5% error values were achieved in the modeling and inverse processes for all five subjects. The average reproduction performance error for the five subjects was about 2.4%. When the equalization filter was absent, on the other hand, the reproduction performance error reached as high as 80%.

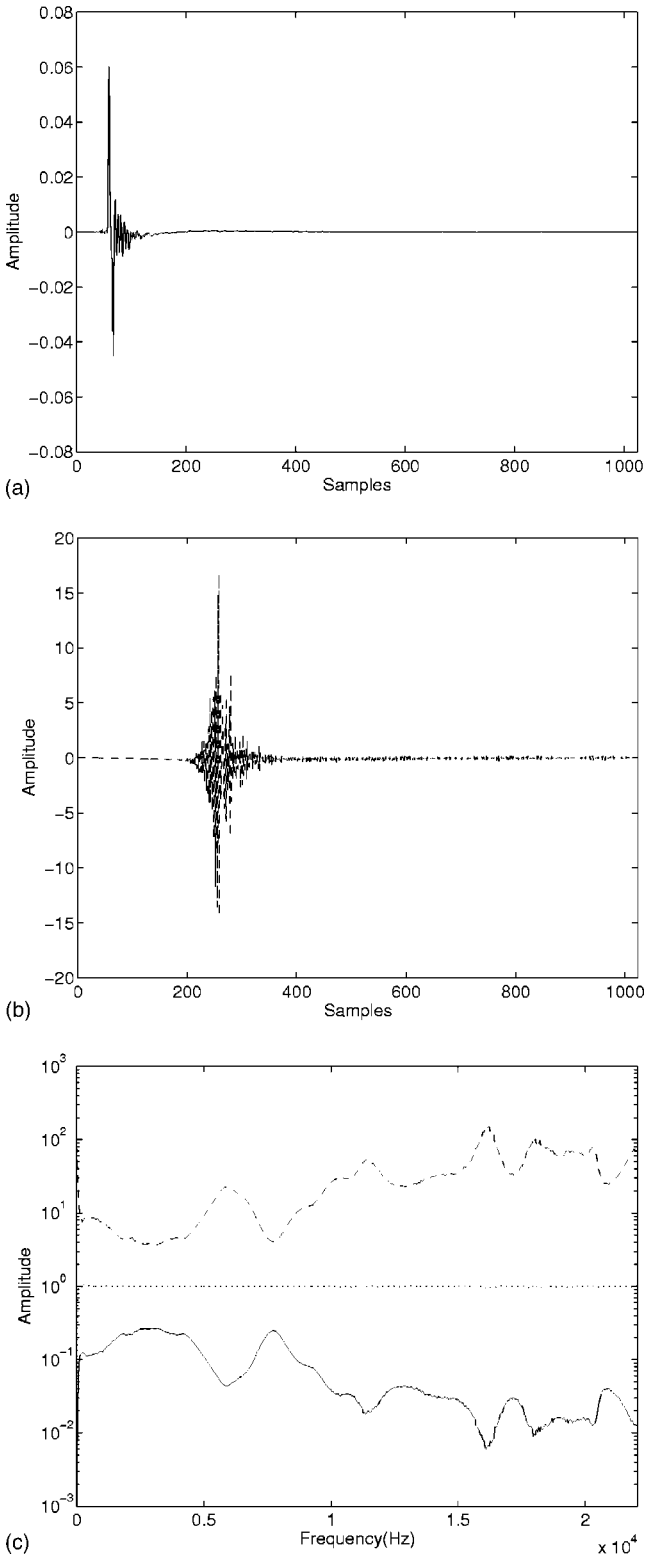


FIG. 4. Headphone-ear canal impulse response $g_1(t)$ of subject W in (a), its inverse $h_1(t)$ (b), and their frequency responses in (c): system model $G_1(j\omega)$ (solid line), its equalization filter $H_1(j\omega)$ (dashed), and the complete channel response (dotted) computed by $G_1(j\omega)\,H_1(j\omega)$.

IV. SUBJECTIVE TEST

A. Test sources and sets

Subjective tests were carried out for each subject immediately after equalization, and were completed for all five subjects. The probe microphone used in the equalization session was pulled out of the ear canal to avoid any discomfort. The original source signal was a random (white noise) signal of 250 ms duration with 20 ms ramps at the beginning and end, respectively. By convolving the source signal with a pair of HRTFs, $A_1(j\omega)$ and $A_2(j\omega)$ in Fig. 1, for a specific source, a pair of virtual source signals were synthesized. The virtual sources synthesized and tested were those in the horizontal plane, equally distributed in 15° steps and distanced 1.4 m from the center of the subject's head. In total, 24 positions were tested starting from the right front and shifting in a clockwise direction, as illustrated in Fig. 5.

Four sets of tests were conducted separately: (1) KEMAR HRIRs without headphone equalization; (2) KEMAR HRIRs with equalization; (3) individual HRIRs without equalization; and (4) individual HRIRs with equalization. KEMAR (nonindividual) HRIRs were those provided by MIT using a KEMAR dummy head (Gardner and Martin, 1994), and the individual HRIRs of the five subjects were measured in an anechoic chamber, as summarized in the Appendix. No information about the test sets (i.e., which set was playing) was given to the subjects. Although two of the subjects (the authors, S and W) were well informed about how the signals were processed, this was not thought to affect the confidence of the tests. It was because the tests were to *identify* the locations of auditory events (i.e., images of sound sources) that had no fixed correct answers.

TABLE I. Modeling, inverse, and reproduction performance errors for each subject (%).

Subjects	Errors		
	Modeling	Inverse	Reproduction
S	0.20	0.33	1.95
W	0.20	0.15	1.75
M	0.18	0.22	3.34
G	0.18	0.44	1.09
D	0.19	0.19	3.85
Average	0.19	0.26	2.40

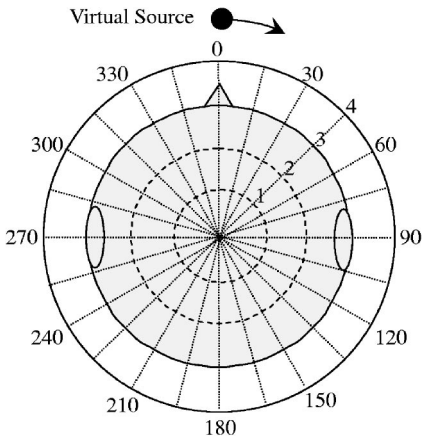


FIG. 5. Test sheet distributed to the five subjects to indicate the image locus of a virtual source starting from the right front direction and shifting in the clockwise direction in 15° steps. At front–back confusion angles, the perceived distance is indicated in the known original direction with the same length but with a mark for distinction.

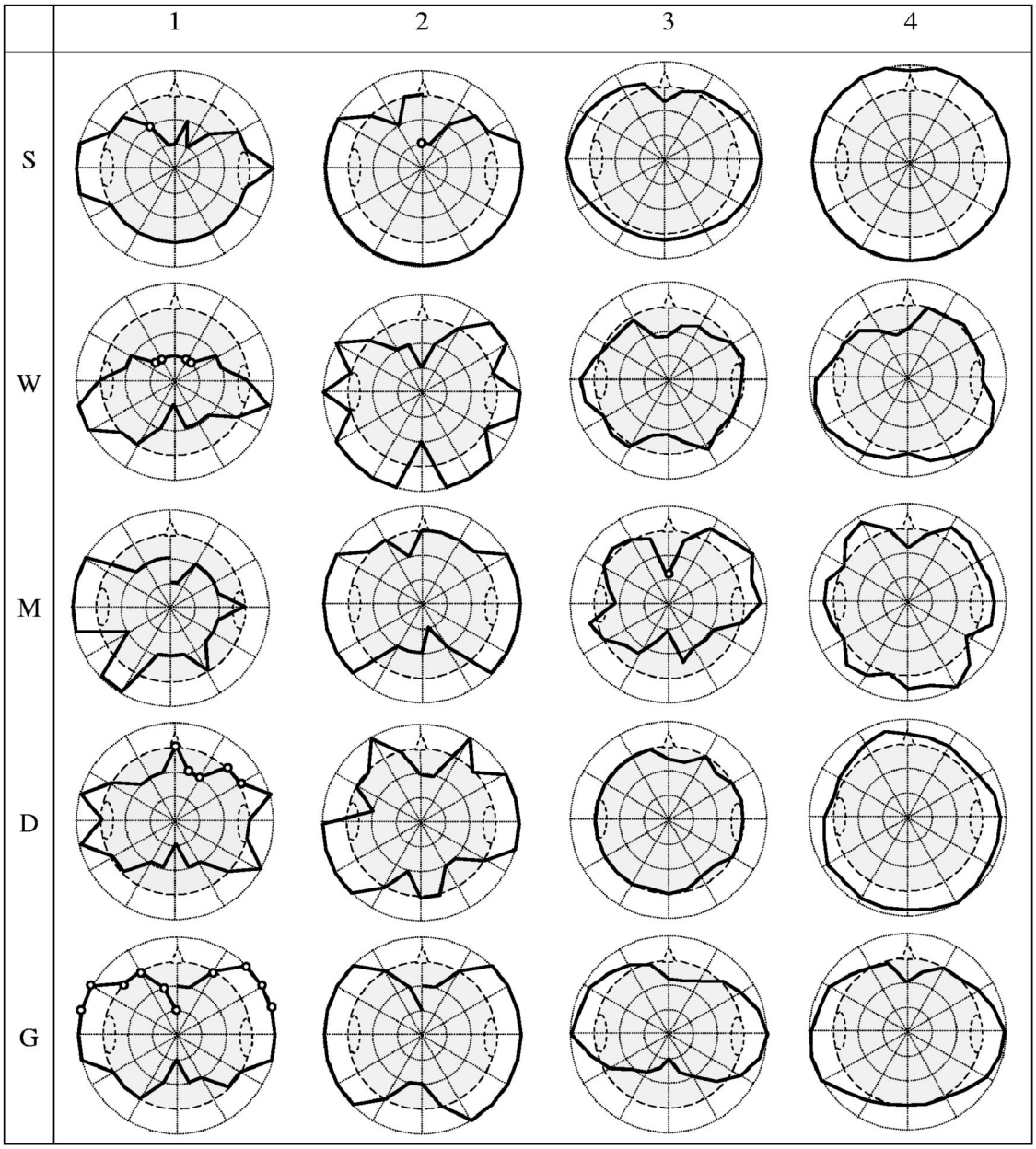


FIG. 6. Source image loci (thick solid lines) identified by subjects S, W, M, D, and G for four test sets: (1) KEMAR HRIRs without headphone equalization, (2) KEMAR HRIRs with equalization, (3) individual HRIRs without equalization, and (4) individual HRIRs with equalization. The small circles indicate front–back confused angles.

B. Instructions

Unlike some subjective tests in the literature (Møller *et al.*, 1996, and references therein), *estimating* a number of ''predefined'' locations of acoustic sources was not a concern in this test. Instead, the tests were focused on a more fundamental problem: the locations of sound images where the auditory events were perceived to occur. In particular, the degree of externalization was investigated in terms of the distance of sound images from the head center, and the direction was disregarded. The direction was excluded because some preliminary tests conducted suggested that a confident determination of both the distance and direction of a source image at the same time was an extremely difficult and stressful task, particularly when the source image was localized inside the head.

Thus in the actual test, each subject was informed of the source direction prior to testing each source, and was instructed to simply indicate the source image distance only along the given source direction. Subjects were also instructed to indicate front–back confusion sources. They were advised to pay careful attention to the relative location of source images with reference to the boundary of the head. They were also advised to hold their heads upright and close their eyes to maximize concentration while sources were replayed, but not forced to do so. Before the actual test, sounds for the set of 24 sources in Fig. 5 were played in a clockwise direction, three times, to give familiarity of the rough locus of the source set. In the actual test, sounds for each source were repeatedly played until the subject made a decision. It was reminded that there was no fixed correct answer; the

location marked by the subject was the answer. Each test set took less than 10 min for each subject.

In Fig. 5, there are three divisions inside the head and one division outside in the test sheet distributed to each subject for indicating the source images using a pen. Such divisions were made since our initial concern was to check whether the sound images of each test set were located inside or outside of the head. Test sets 1 and 2 were conducted first, and sets 3 and 4 were tested several weeks later. The former sets were marked on the discrete divisions, 0 to 4. On this scale, the images near the vicinity of the skull are marked on level 3 and those ''clearly externalized and remote'' are marked on level 4. After conducting sets 1 and 2, however, some subjects suggested that a continuous scale would be more appropriate (perhaps regarding the inside in linear scale and the outside in logarithmic scale). As such, this scale was adopted for sets 3 and 4 without a serious loss of consistency. Again, level 4 is for ''clearly externalized remote'' images and no specific distance for it was indicated to the subjects.

C. Discussions

The results from the five subjects are combined for all four test sets in Fig. 6 for easy comparison, where the columns correspond to test sets 1–4 counting from the left while the rows correspond to subjects S, W, M, D, and G from the top. The thick solid lines indicate the perceived sound image loci and the small circles in the loci denote the front–back confusion angles. For example, in the (1,1) subplot showing the test result of subject S for set 1, the actual source from 330° was perceived to be from 210°, and its absolute image distance was level 2, according to the test sheet in Fig. 5.

Set 1 (KEMAR HRIRs without equalization) seriously suffers from in-head-localization in most angles as well as front–back confusion. All the confusions are front-to-back confusions; that is, front sources are perceived to be rear sources. Only the right and left sources at around 90° and 270° are perceived to be externalized. Set 2 (KEMAR HRIRs with equalization) offers dramatically improved spatial effects to all subjects. Front–back confusion is greatly reduced and many sources are now externalized except the front and rear sources within the scope angle of ±30°. It is shown that the front sources are more difficult to externalize in subjects S and W, and the front and rear sources are equally difficult in subjects M, D, and G. A comparison of the results of sets 1 and 2 makes the advantage of headphone equalization obvious: it enlarges the source image distance so that, as a benefit, externalization as well as front–back distinction is enhanced.

Since the imperfect results in set 2 were suspected to be from the dummy head HRIRs, which could be quite different from those of the subjects, each individual's HRIRs were used in sets 3 and 4. Test results of set 3 (individual HRIRs without equalization) show that front–back distinction is improved and the image distance is enlarged, compared with set 1. In addition, the most remarkable pattern is that the distance perception is now more consistent and the loci are more circle-like, as observed particularly in subjects, S, W, and D. On the contrary, there has been a lot of fluctuation in the loci of sets 1 and 2 using the KEMAR HRIRs. Comparing set 2 and set 3, it can be noticed that the use of equalization is more important for externalization than the use of the individual HRIRs. Finally, results of test set 4 (individual HRIRs with equalization) are shown in the fourth column. No front–back confusion occurs in any of the subjects. Subjects S and D report perfect externalizations with consistent distances, subject M reports an excellent externalization with a moderately consistent distance, and subject G reports an improved spatial perception compared with set 2. In subject W, sources on the front half are not externalized for some reason, but distance perception is slightly more consistent than set 2.

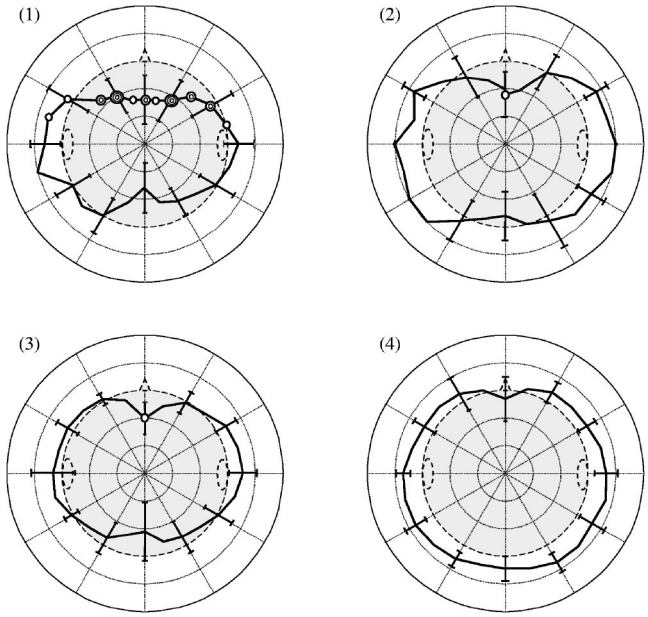


FIG. 7. Average source image loci (thick solid lines) and standard deviations (error bars) for four test sets 1, 2, 3, and 4 are shown in the subplots (1)–(4). The number of concentric circles in the loci denotes the number of front–back confusions among five subjects. For detailed captions on each set, refer to Fig. 6.

Figure 7 concisely compares some statistical values of each test set from the raw data shown in Fig. 6. The subplots of Fig. 7 correspond to the four test sets, as indicated by (1)–(4). The mean (thick solid lines) and standard deviation (error bars) values were calculated at each angle considering only the absolute image distances. Front–back confusions are separately considered using concentric circles, where the number of concentric circles indicates the number of occurrences from five subjects. It should be remembered that the statistical evaluations in Fig. 7 are based on data from only five subjects that are rather small samples to draw any statistical generalization. Thus, any general discussion on the trend should be made by cross-checking the consistency with the original data in Fig. 6. The well externalized source directions in each test set are the side sources at around 90° and 270° in set 1, the lateral sources within the scope angle ±45° in set 2, and the lateral sources within the scope angle ±30° in set 3, and nearly all the surrounding sources in set 4. Sets 1–3 have ''cocoon-like'' shapes with large blunt notches up and down, while set 4 has nearly a ''circle'' shape with a small notch at the top.

The cocoon-shape results could be due to the variable human resolution to horizontal spatial sources. The localization blur (perceptual resolution or sensitivity) of humans has a global minimum of $\pm 3.6°$ for the right front source, a local minimum of $\pm 5.5°$ for the rear, and has increasing blurs as the source moves to the sides (Blauert, 1983, pp. 40–41). Thus, humans are more sensitive in discriminating frontal and rear sources than lateral sources. It may equally mean that humans are not easily fooled (convinced) by frontal or rear virtual sources, unless the virtual sounds are very close (unperceivable) duplications of the corresponding real sounds. For lateral sources, to which humans are less sensitive, a crude virtual sound system such as using test set 1 can still work well as shown in Fig. 7(a). In addition to this hypothesis, the notch on top of the circle could be partly due to the fact that some other sense, such as vision, might be more crucial for the human perception of the right front source (Toole, 1969). In fact, difficulties in the median plane sources, the right front and rear sources, are occasionally experienced even in real-life listening (Toole, 1969; Møller *et al.*, 1996).

Overall, it was observed that, when the image was localized inside the head, more playbacks of the signals were requested and the decision often turned out to be a front–back confusion angle. The right front source at 0° was the most repeated one for all subjects in all test sets. Set 1 took the longest time for testing, sets 2 and 3 followed sequentially, and set 4 was the shortest. To account for these differences, it may be useful to introduce a confidence level for each decision, indicated in parallel with the image location. Or as an alternative, the test may be performed several times. Finally, it is worthwhile to add some *post hoc* discussions on the subjective test sheet in Fig. 5, where there are three equally distanced divisions inside the head and one division outside. The divisions can be made several different ways including the suggestion of one of the reviewers of this paper: the apparent distance is simply ''about right'' or ''too close'' or ''inside.'' If most sources are imaged inside the head, such as the case of test set 1, many divisions outside the head extending to the presumed distance (1.4 m) or the auditory horizon would be of no use, but distracting the subject who is paying attention to the relative distance of the image with respect to his skull. However, this choice of divisions can be helpful when most sources are imaged outside, such as the case of test set 4. Thus, the divisions may have to be made in an ad hoc manner depending on applications.

V. CONCLUSIONS

The degree of externalization in headphone reproduction was investigated for various synthesis and reproduction cases. An effective binaural headphone system was constructed by way of binaural synthesis using head-related impulse responses and individual headphone equalization using Wiener filter theory. The headphone system designed had an average reproduction performance error of 2.4% for five subjects with a random noise input, and was used to perform some subjective tests for a set of virtual sources in the horizontal plane which were equally distanced 1.4 m from the center of each subject's head.

In subjective tests, the effects of individual and nonindividual binaural syntheses and those of equalized and nonequalized reproductions were separately investigated. Although the number of subjects who participated was only five, the data across the subjects were quite consistent to reveal some trends. The well externalized source directions obtained in each case were: the side sources at around 90° and 270°, when nonindividual synthesis was used without equalization (set 1); the lateral sources within the scope angle $\pm 45°$, when nonindividual synthesis was used with equalization (set 2); the lateral sources within the scope angle $\pm 30°$, when individual synthesis were used without equalization (set 3); and nearly all the surrounding sources except the right front source, when individual synthesis were used with equalization (set 4). The results also demonstrate that individual equalization (sets 2 and 4) is important for externalization, and individual synthesis (sets 3 and 4) is important for consistent distance perception. Thus, a combined use (set 4) of both individual equalization and individual synthesis resulted in externalized sound images with a reasonably consistent distance. With the combined use, it may now be possible to make a simultaneous confident judgment on both the distance and direction of auditory events as well as those of acoustic sources.

ACKNOWLEDGMENT

This work was supported by Grant No. R05-2003-000-11999-0 from the Korea Science and Engineering Foundation.

APPENDIX: HRIR MEASUREMENT

Head-related-impulse-responses (HRIRs) of the five subjects were measured in an anechoic chamber (Choi, 2004). The sound sources tested were located in 24 directions equally spaced in the horizontal plane at a distance of 1.4 m from the center of the subject's head. Figure 8 shows the test setup, where a swiveling hairdresser's chair has been modified for the test. The subject, who was seated in the chair with a homemade headrest embracing up to just behind the ears, was instructed by an operator outside the chamber to change his orientation with respect to the loudspeaker direction using the mirror reflecting the protractor marked on the round table. Just above the protractor, a notch mark was welded to the rotating part of the chair to help a correct angle setting. The use of a pivotal mirror holder and a height-adjustable headrest was helpful to accommodate many subjects with different heights. They could easily adjust the correct angle by slightly looking down without tilting their heads. A MSP5 Yamaha monitor speaker was used, two Etymothic Research ER-7C probe microphones were inserted about 20 mm into both ear canals of each subject, and the measurement system was controlled by a PC using the MATLAB data acquisition and signal processing toolboxes with a LYNX II soundcard.

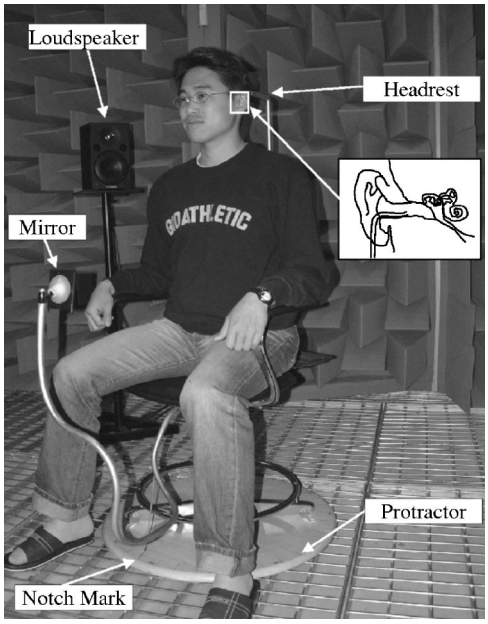


FIG. 8. HRIR measurement setup. Being seated in a swiveling hairdresser's chair with a homemade headrest embracing up to just behind the ears, the subject was instructed by an operator outside the chamber to change his orientation with respect to the loudspeaker direction using the mirror reflecting the protractor marked on the round table. Just above the protractor, a notch mark was welded to the rotating part of the chair to help a correct angle setting. Two probe microphones were inserted into both ears of the subject.

The test procedure closely followed that of Gardner and Martin (1994). Thus, a MLS of length $2^{14}-1$ was used to accelerate the system identification process as discussed in Sec. II, the number of averages was five, and the sampling frequency was 44.1 kHz. The modeling performance errors were mostly less than 0.4%, but reached up to 4% when the measurement was made in the head shadow region. In order to exclude the transducer effects in the raw HRIRs, an additional test was also conducted by removing the subject and placing the microphones at the original head center position. However, the data processing method to yield a final HRIR here is slightly different from those in the literature. Let $g_{\text{raw}}(t)$ and $g_{\text{free}}(t)$ denote the HRIRs with and without the subject in place, respectively, and whose frequency domain counterparts are $G_{\text{raw}}(j\omega)=S(j\omega)A_{\text{raw}}(j\omega)M(j\omega)$ and $G_{\text{free}}(j\omega)=S(j\omega)A_{\text{free}}(j\omega)M(j\omega)$, respectively. Here, $S(j\omega)$ and $M(j\omega)$ are, respectively, the frequency response functions of the loudspeaker and microphone used, and $A_{\text{raw}}(j\omega)$ and $A_{\text{free}}(j\omega)$ are those of the acoustic systems with and without the subject, respectively. The HRTF, which is the frequency domain representation of a HRIR, may be written as $A(j\omega)=G_{\text{raw}}(j\omega)/G_{\text{free}}(j\omega)$. However, such a simple calculation in the frequency domain generates serious stability and causality issues as discussed by Kim and Wang (2003). As can be noticed from $A(j\omega)=G_{\text{raw}}(j\omega)/G_{\text{free}}(j\omega)$, this is to design a physically realizable filter $A(j\omega)$, which can be tackled by the Wiener filter explained in Sec. II. Thus, the optimal HRIR can be obtained using a deterministic Wiener filtering by setting the received signal to $r(t)=g_{\text{free}}(t)$ and desired signal to $d(t)=g_{\text{raw}}(t-\tau)$, where τ should be sufficiently large to make the HRIR causal. This method is particularly effective when $g_{\text{raw}}(t)$ and $g_{\text{free}}(t)$ have short reverberation times such as those measured in an anechoic chamber. Otherwise, an approximation method can alternatively be used, as explained in Sec. II.

Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization*, Revised Edition (MIT, Cambridge, MA).
Choi, W. J. (**2004**). "Analytical and experimental modeling of HRTFs for the authentic reproduction of 3D sound," M.Sc. dissertation, Gwangju Institute of Science and Technology, Korea.
Elliott, S. (**2001**). *Signal Processing for Active Control* (Academic, London).
Gardner, B., and Martin, M. (**1994**). "HRTF measurements of a KEMAR dummy-head Microphone" (MIT's Media Lab, http://sound.media.mit.edu/KEMAR.html).
Hartmann, W.M., and Wittenberg, A. (**1996**). "On the externalization of auditory images," J. Acoust. Soc. Am. **99**, 3678–3688.
Haykin, S. (**1996**). *Adaptive Filter Theory*, 3rd. ed. (Prentice-Hall, Englewood Cliffs, NJ).
Kim, S.M., and Wang, S. (**2003**). "A Wiener filter approach to the binaural reproduction of stereo sound," J. Acoust. Soc. Am. **114**, 3179–3188.
Kirkeby, O., Nelson, P.A., Hamada, H., and Orduna-Bustamante, F. (**1998**). "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**, 189–194.
Møller, H., Sørensen, M.F., Jensen, C.B., and Hammershøi, A.D. (**1996**). "Binaural techniques: Do we need individual recordings?," J. Audio Eng. Soc. **44**, 451–469.
Oppenheim, A.V., and Schafer, R.W. (**1989**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Papoulis, (**1984**). *Signal Analysis* (McGraw-Hill, New York).
Sakamoto, N., Gotoh, T., and Kimura, Y. (**1976**). "On out-of-head localization in headphone listening," J. Audio Eng. Soc. **24**, 710–716.
Toole, F.E. (**1969**). "In-head localization of acoustic images," J. Acoust. Soc. Am. **48**, 943–949.
Vanderkooy, J. (**1994**). "Aspects of MLS measuring systems," J. Audio Eng. Soc. **42**, 219–231.
Wightman, F.L., and Kistler, D.J. (**1989a**). "Headphone simulation of free field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Wightman, F.L., and Kistler, D.J. (**1989b**). "Headphone simulation of free field listening. II. Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.
Zahorik, P. (**2000**). "Limitations in using Golay codes for head-related transfer function measurement," J. Acoust. Soc. Am. **107**, 1793–1796.

The vibro-acoustic response and analysis of a full-scale aircraft fuselage section for interior noise reduction

Peter C. Herdic,[a] Brian H. Houston, Martin H. Marcus, and Earl G. Williams
Naval Research Lab, Physical Acoustic Branch, Washington, DC 20375

Amr M. Baz
University of Maryland, College Park, Maryland 20742

(Received 12 August 2004; revised 21 December 2004; accepted 17 February 2005)

The surface and interior response of a Cessna Citation fuselage section under three different forcing functions (10–1000 Hz) is evaluated through spatially dense scanning measurements. Spatial Fourier analysis reveals that a point force applied to the stiffener grid provides a rich wavenumber response over a broad frequency range. The surface motion data show global structural modes ($\lesssim$150 Hz), superposition of global and local intrapanel responses ($\sim$150–450 Hz), and intrapanel motion alone ($\gtrsim$450 Hz). Some evidence of Bloch wave motion is observed, revealing classical stop/pass bands associated with stiffener periodicity. The interior response ($\lesssim$150 Hz) is dominated by global structural modes that force the interior cavity. Local intrapanel responses ($\gtrsim$150 Hz) of the fuselage provide a broadband volume velocity source that strongly excites a high density of interior modes. Mode coupling between the structural response and the interior modes appears to be negligible due to a lack of frequency proximity and mismatches in the spatial distribution. A high degree-of-freedom finite element model of the fuselage section was developed as a predictive tool. The calculated response is in good agreement with the experimental result, yielding a general model development methodology for accurate prediction of structures with moderate to high complexity.
 [DOI: 10.1121/1.1887125]

PACS numbers: 43.40.Rj, 43.40.Ey, 43.40.Yq [SFW] Pages: 3667–3678

I. INTRODUCTION

Interior sound radiation in aircraft cabins occurs through a number of transmission paths and vibro-acoustic mechanisms. It is well known that at frequencies below approximately 400 Hz high interior acoustic levels are generated mainly by the propulsion system of the aircraft and transmitted through both structure-borne and air-borne paths. For example, a dominant type of structure-borne noise originates through rotating unbalance of the aircraft engine.[1–3] The associated vibrational motion travels through the structure to the fuselage walls and radiates into the interior. Air-borne noise transmission paths can involve engine exhaust induced vibrations[4] or propeller generated pressures on the fuselage skin.[5–7] Other noise sources like propulsion harmonics and gearbox noise may be present, but they are lower in level and closer to the flow-induced noise floor. This paper will examine the response and associated vibro-acoustic mechanisms of a Cessna fuselage section in a laboratory environment. Three excitations, that will be examined separately, include point forces applied at two locations on the fuselage and an external acoustic loudspeaker. The physics of the response will be analyzed toward the goal of reducing the interior noise, but the actual implementation of sound quieting approaches will be in a future paper.

Cheng,[8] Pope *et al.*,[9] and others have analytically examined the structural and interior acoustic response of a simple end-plated cylindrical shell subjected to point excitation. The approach in these papers is to examine the coupled vibro-acoustic response as a result of interactions between the cylindrical shell, the flat end plates, and the interior cavity. Normal cavity modes form on the interior associated with the rigid wall geometry of the cylindrical shell. Often the structural modes associated with the cylindrical shell and the ends force their response on the interior, even where no cavity modes exist. However, these modes often do not couple well to the interior due to a mismatch in the spatial distribution between the structural mode shapes and interior normal cavity modes. Snyder and Hansen[10] showed that coupling of the structural and acoustic modes only occurs with an exact match in the circumferential mode number and odd/even matching of axial mode number. Fahy[11] and Henry *et al.*[12] both discuss mode matching with the requirement of frequency proximity. For mode coupling to occur under the frequency proximity criteria, the structural mode must have sufficient damping to span frequencies to include at least a portion of a damped interior acoustic mode.

The vibro-acoustic response in fuselage structures is more complex than idealized by simple end-plated cylindrical shells due, in part, to the structural discontinuities, disorder and the pseudoperiodic nature associated with the ribs, stringers and the other internal structures. The response due to the presence of exactly periodic discontinuities has been studied in ribbed plates and cylindrical shells.[13–15] As a flexural wave encounters a discontinuity, a portion of the wave is transmitted past while another part is reflected. The reflected portion from the multiple exactly periodic discontinuities is coherent, distributing energy only in certain frequency bands. This distribution creates allowed passbands and disal-

[a]Also with SFA Inc., Largo, MD 20775; electronic mail: herdic@code7136.nrl.navy.mil

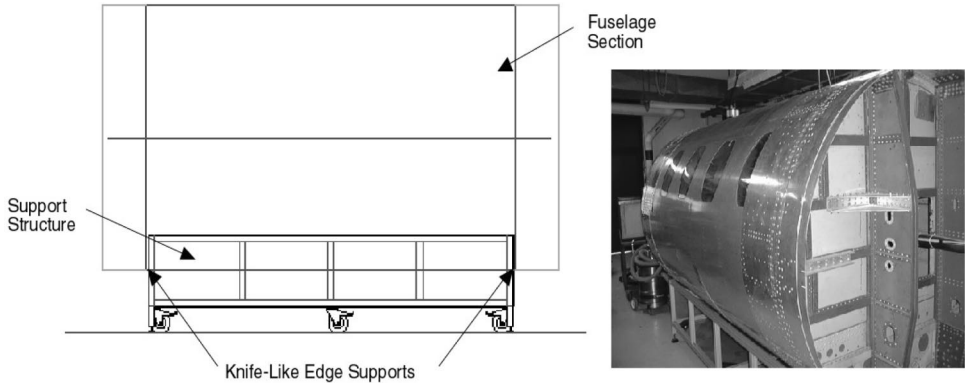


FIG. 1. Fuselage section and support structure.

lowed stopbands. In solid-state physics, this motion is commonly referred to as Bloch waves,[16] after F. Bloch who first identified them in lattice structures with exactly periodic atoms. Photiadis[17] notes that this motion also follows Floquet's theorem,[18] which asserts a traveling wave solution that is the consequence of translation invariance by multiples of the rib spacing. In addition, complex structures also exhibit confinement or localization of energy in small regions of space due to variations in periodicity. Hodges[19] first recognized the Anderson localization phenomenon[20] in dynamical systems when exploring the behavior of an array of coupled pendula with disorder. Anderson localization has also been observed in structural systems such as plates and cylindrical shells with rib variability.[21–23] Some evidence of Bloch waves is present in these fuselage section measurements, however, Anderson localization is more difficult to identify and may be a subject of a future paper.

Complex fuselage structures have been modeled numerically using finite element approaches by a number of researchers[7,24–27] using various codes. The models tend to include a high level of detail with a very large number of degrees-of-freedom. The models can be on the order of 1 000 000 degrees-of-freedom for larger structures where the responses are calculated below 400 Hz. Evidently, this model size is required to accurately predict modal frequencies with average errors less than 10% when compared with experimental results.[25,26] The model developmental methodology illustrated in this paper will result in an accurate prediction of not only mode frequencies, but also the absolute level and distribution of the spatial response. The numeric model size is smaller since only a fuselage section is examined, but the level of model detail remains approximately the same.

II. MEASUREMENT APPARATUS

The fuselage section response measurements were performed at the Laboratory for Structural Acoustics at the Naval Research Laboratory, Washington, DC. Typical measurements made in the Air Acoustics Experimental Facility involve a scanning robot to provide response data with high spatial sampling. This robot can be instrumented with a number of different transducers. In the measurements described in this paper, the automated scans were performed with a six element microphone array and an out-of-plane Laser Doppler Vibrometer (LDV) probe to measuring interior pressure and surface velocity, respectively.

The experiments were carried out on a Cessna Citation II fuselage section. The fuselage section configuration may be described generally as an aluminum stiffened cylindrical shell held in a simply supported-like manner (see Fig. 1). Two knife-like edges (over 45° of the circumference) support the section at either end. The support edges were positioned directly below the two flat bulkheads that terminate the section. One bulkhead has a 46 cm by 46 cm square opening for insertion of the robot scanner arm. The fuselage section's interior dimensions consisted of a 0.81 m radius and a 2.55 m length, and the shell thickness varies between 0.8 and 1.2 mm. It can be seen in Fig. 2 that a moderate to high degree of internal structural complexity is present, including pseudoperiodic spaced ribs (9 along floor and 19 along ceiling) and stringers (22). These internal stiffeners become discontinuous in the area of the window sections, and they vary greatly in terms of size and shape. For example, significantly enlarged stringers were present directly below the windows and the area moments of inertia of some ribs varied by as much as 90% around the circumference.

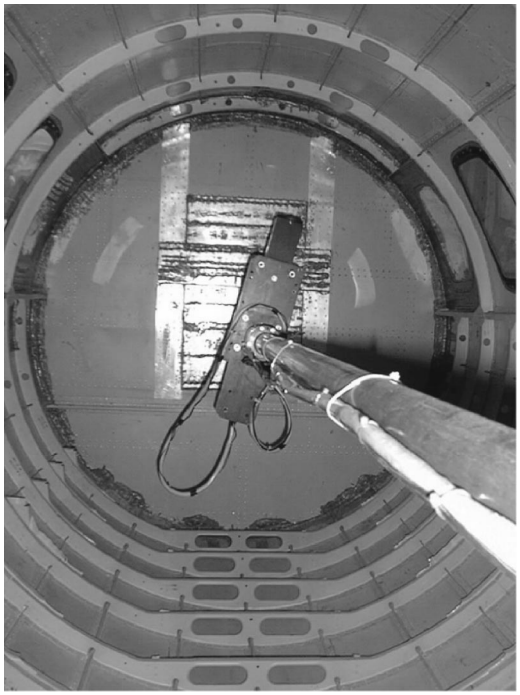


FIG. 2. Internal fuselage complexity and Laser Doppler Vibrometer (LDV) probe attached to automated robotic scanner.

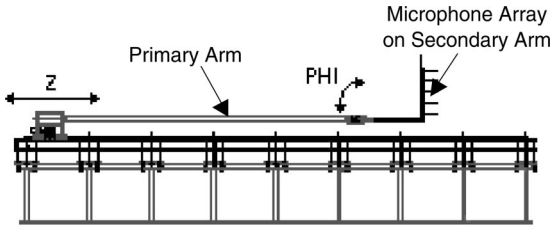


FIG. 3. Cylindrical coordinate based robotic scanner instrumented with microphone array.

Some of the notable additional complexity includes five Plexiglas™ window sections present on each side of the fuselage and portions of seating mounts in the floor area. Details such as an interior trim, insulation, flooring, carpet, seats, etc., were not present.

The automated scans were performed using a cylindrical coordinate (r,z,ϕ) based electro-mechanical robot (Fig. 3). For the pressure measurements, the radial degree of freedom was fixed with a 12.81 cm spacing of 6 microphones along the length of the secondary arm. This arm was designed to maximize the structural stiffness while minimizing the acoustic scattering cross section. The axial and azimuthal degrees of freedom were automated by a drive train and low noise microstepping motors. During the execution of a scan, the scanner arm was inserted in the open end of the fuselage and extended to the far end. Data were acquired by each microphone in equal angular increments, then stepped incrementally in the axial direction and rotated again. This process was then repeated down the length of the fuselage to complete the data set. The microphone array was replaced by the LDV probe to make the out-of-plane surface velocity scan measurements in a similar manner. The LDV probe was attached to the primary arm of the robot as shown inside the fuselage in Fig. 2.

III. VIBRO-ACOUSTIC RESPONSE

The structural response of the fuselage and the associated interior acoustics were examined experimentally under three separate excitation conditions: (1) a point force applied at a rib/stringer intersection at one end of the fuselage, (2) a point force applied to the flexible thin walled panel area between the stiffeners at one end of the fuselage, and (3) an external acoustic source 1 m away from the fuselage skin at top center. The measurements were conducted using a chirp waveform over a band from 10 to 1000 Hz with 0.61 Hz resolution. The spatial density of the scan measurement points varied somewhat, but the nominal size of the data matrices was 30 axial by 64 angular ($\sim$1900 data points) for the surface velocity measurements and 29 axial by 64 angular by 3 radial ($\sim$5500 data points) for the interior pressure.

A. Structural response

The rms (root-mean-squared) spatial average of the scanned surface velocity response measurements is shown in Fig. 4 for the three separate excitation cases. The rms response to the point force applied to the stiffener intersection (case 1) appears highly reactive, exciting a strong structural response over a broad frequency range. The point force applied to the flexible panel (case 2) strongly excites the global structural modes at low frequency, but the higher frequency response appears weaker. The drive point impedance for these two cases is shown in Fig. 5. The soft impedance of the case 2 drive condition clearly regulates the response of the entire structure. An almost one-to-one correspondence can be found when examining the inverted frequency-dependent shape of the impedance and rms response curves. At higher frequencies, the fuselage motion associated with case 2 tends to localize with little propagation past the area of the driven panel. The strong response due to the case 1 excitation is a result of driving the stiffener grid directly, where drive motion is easily transmitted throughout the structure, exciting many modes.

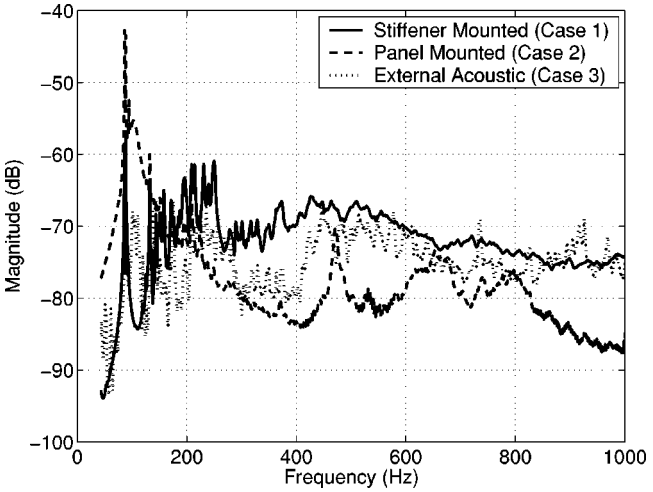


FIG. 4. rms normal surface velocity due to three excitation cases. A point force [(m/s)/N] applied at a rib/stringer stiffener intersection (case 1), a point force [(m/s)/N] applied at a panel (case 2), and external acoustic [(m/s)/Pa] excitation (case 3).

The external acoustic source is a backed electromagnetic woofer with a beamwidth that varies from $\sim$360° (100 Hz) to $\sim$90° (1 kHz) in both vertical and horizontal dimensions. Preliminary measurements of the woofer surface show that it does not break up into any complex modes below 1 kHz, and

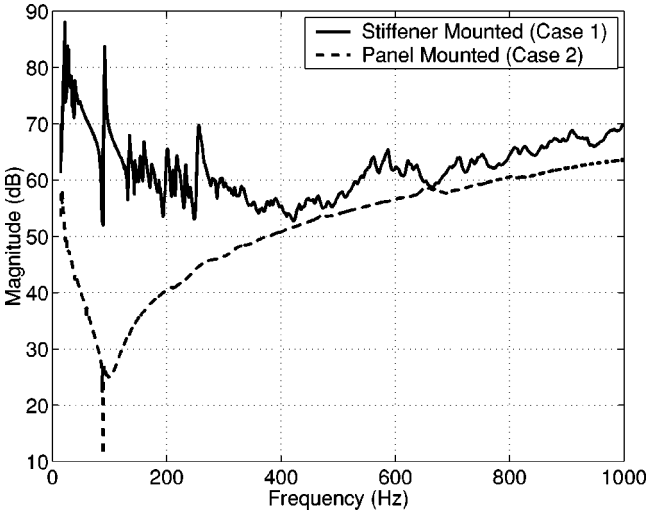


FIG. 5. Drive point impedance [N/(m/s)] due to a point force applied to a stiffener intersection (case 1) and applied to the flexible panel fuselage wall (case 2).

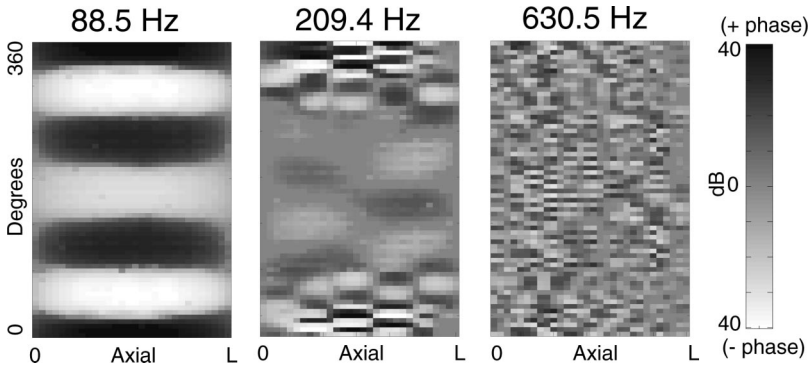


FIG. 6. Phase display of surface velocity illustrating the response categories in the low, mid, and high frequency bands. The light and dark colors indicate surface motion that is 180° out of phase. The categories are illustrated here using a point force excitation applied at the stiffener intersection (case 1).

therefore, the acoustic radiation impedance approximately follows that of a baffled circular piston. An average sound pressure level of 77 dB (50–1000 kHz) was measured at a distance of 0.5 m from the woofer. The room where the experiments were conducted is highly structured in an asymmetric manner so that no strong room modes existed. The rms response associated with the external acoustic source, case 3, demonstrates a fairly high reactive response over a broad frequency range, but it is somewhat weaker below 400 Hz. Note that the response associated with the point force excitations and external acoustic source have different units; it is only coincidental that the curves lie near the same magnitude level. Since the focus of this paper is on the low frequency response and the clear excitation of the dominant vibro-acoustic mechanisms within this regime, the case 1 excitation appears most relevant. Therefore, the point force excitation applied to the stiffener intersection will be used in much of the ensuing analysis in this paper.

The structural response of the fuselage section can be broken up into three general categories in specific frequency regimes. These categories are illustrated using the real space (circumferential angle versus axial position) normal surface velocity displays shown in Fig. 6. The displays are derived for the stiffener point driven case (case 1), however, the other excitations exhibit a similar categorization to slightly varying degrees. The plots are phase compensated where the real part of the phase-shifted response shows the positive (dark areas) and negative (light areas) directions of the motion. The first category consists of global modes at low frequency ($\lesssim$150 Hz). The display of normal surface velocity at 88.5 Hz illustrates an $m=1$, $n=3$ global mode that is typical in this response category, where n refers to circumferential mode numbers and m to axial mode numbers. The only other distinct global response is a predominantly $n=4$, $m=1$ mode at 132.4 Hz (not shown). The second category involves a transition region that occurs at mid-frequencies ($\sim$150–450 Hz). In this band, strong local structural responses are superimposed on lower amplitude global structural modes. For example, the response at 209.4 Hz shows high local levels on the crown of the fuselage with a lower level global $m=\sim 2$, $n=\sim 4$ component. The local intrapanel responses are due to low order modes of the individual panels of the fuselage. The third category represents an entirely intrapanel response. Multiple half structural wavelengths fill the sub-bay areas divided by the ribs and stringers. The surface velocity response at 630.5 Hz is typical at these higher frequencies ($\sim$450–1000 Hz). The intrapanel response is somewhat hard to observe in this display since the spatial sampling is beginning to approach the Nyquist sampling criteria for this measurement. Figure 7 shows the phase response display in a single bay with much higher spatial resolution. Three circumferential half-wavelengths and one axial half-wavelength can be observed in the single bay between the frames and stringers. These intrapanel responses are simply modes of the individual panels, which could be predicted by plate theory if the complicated boundary conditions could be understood. Some experimental results exist[28] for riveted curved cylindrical panels where it is noted that there is little difference in the natural frequencies between flat and curved panels when the node lines are parallel to the longitudinal direction. The curvature effects stiffen the panel in the axial direction which increase the axial mode frequencies and leave relatively high circumferential mode numbers at low frequency.

Several sets of local response data (case 1) over specific regions of the fuselage are shown in Fig. 8. The two mosaics on the left contain surface velocity data on a line along the length of the fuselage and a ring around the circumference displayed in real space position versus frequency. At high frequencies, the energy is clearly regulated by the stiffener spacing. The displays on the right are the associated one-dimensional spatial Fourier mode decompositions. Clear banding associated with Bloch wave motion is seen where energy is allowed (passband) and disallowed (stopband) in certain frequency regions. Banding transitions should occur at each Brillouin zone[18] with $k=\pi/a$ or $m=L/a$ intervals

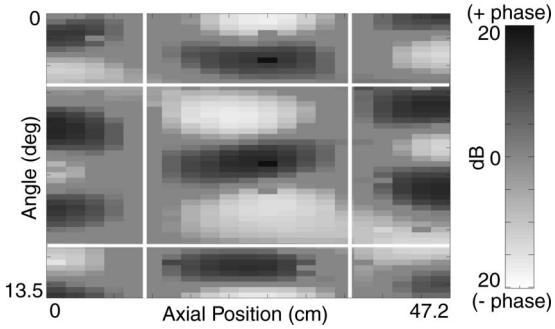


FIG. 7. Surface velocity phase response at 630.2 Hz with increased spatial resolution. The display shows a single fuselage bay due to a point excitation applied to the stiffener intersection (case 1). The white lines indicate the locations of frames (vertical) and stringers (horizontal).

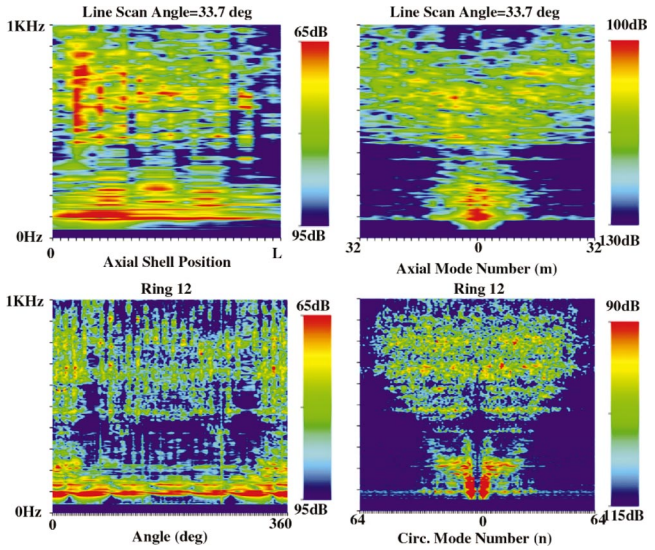


FIG. 8. Surface velocity [(m/s)/N] response in local regions due to a point excitation applied to the stiffener intersection (case 1). The top displays are along a line at 33.7° from crown and the bottom displays are at a ring near the middle of fuselage. Real space displays of the line and ring are shown on the left and the associated wave number decomposition is on the right.

where a is the exact spacing of the periodicity. For this fuselage, the first banding transitions in the line data should occur at $m=\sim 5$ based on the averaged rib spacing on the bottom half of fuselage and at $m=\sim 10$ due to the averaged spacing on the top half. In the ring data, the averaged spacing of the stringers indicates the first banding transition should occur at about $n=\sim 13$. The banding does not occur exactly at these predicted mode numbers for either the line or ring data indicating that the periodic variations and other complexities in the fuselage are modifying the result. The fuselage is psuedoperiodic, whereas, the band transitions are based on exact periodic spacing. The variability in frame spacing alone changes from 23 to 31 cm along the length, significantly breaking the periodicity.

It should be pointed out that this data ring is based on a higher spatial sampling (128 circumferential points) than was used in the two-dimensional experimental scans (64 circumferential points). The spatial sampling for the two-dimensional experimental scans would lead to a maximum circumferential mode number of ± 32, based on spatial Nyquist sampling theory. In viewing the spatial transform of the ring, small amounts of spatial aliasing could begin as low as 500–700 Hz in the complete two-dimensional experimental scans.

The full two-dimensional spatial Fourier decomposition of the normal surface velocity data into wavenumbers is shown in Fig. 9 for the case (1) excitation. These displays show the mode decompositions in terms of axial mode number versus frequency for each circumference mode number up to $n=7$. The strong low frequency $n=3$ response can be observed, as well as the other strong global mode at $n=4$. The dispersion curves become increasingly less distinct, or smeared, at higher frequencies. Clear branches of either compressional or flexural dispersion curves are difficult to discern. These branches are crisp and distinct in numerical calculations where the framed shell model has low to moderate complexity.[24] This smearing is consistent with the complex nature of the fuselage test structure, where pseudoperiodicity and discontinuities play a significant role. The stop/pass banding associated with periodicity is not clearly illustrated here, as compared to that observed in the local regions shown in Fig. 8. The periodicity in the local regions is perhaps more exact, and when the full two-dimensional fuselage is taken into account, the periodicity appears even more random and the banding effects less evident. Also, the relatively high level of damping in the fuselage may play a role in smearing the energy out over the entire band.

The sonic lines shown in the displays of Fig. 9 illustrate the structural wave transition from subsonic to supersonic acoustic radiation. The area enclosed within the cone ($n=0$) or parabolic curves ($n \geqslant 1$) is the supersonic region, and the area exterior to this region is subsonic. Examination of the decompositions shows that all strong surface responses are subsonic. Sound radiation problems are normally thought of in terms of the farfield, where it is the supersonic responses that are strong radiators. The subsonic part decays quickly away from the vibrating surface, never reaching the farfield. These pressures typically decay completely within a short distance from the structure's surface through a hydrodynamic short-circuit mechanism, where the fluid sloshes back and forth between structural responses of alternating phase. As will be discussed in the following, it is important to realize that a portion of the aircraft interior space is within the acoustic nearfield and that both the supersonic and subsonic responses are important to the interior acoustic response.

B. Interior acoustics

The interior acoustic response at low frequency is driven or forced by global structural modes. An example of this mechanism is illustrated in Fig. 10 using the response due to the case (1) excitation. These phase displays show the normal surface velocity and the interior pressure at the 88.5 Hz, $m=1$ $n=3$ mode. For the surface motion displays, the light color indicates a high level of motion in the positive or outward direction and the dark regions are high levels in the negative or inward direction. The light colored regions in the interior pressure displays indicate rarefaction or expansion and the dark regions indicate compression. The interior pressure displays are slices through the center of the fuselage, both axially and circumferentially. The interior pressure and structural response match in spatial distribution and phase, illustrating that the strong $n=3$ $m=1$ structural mode is forced or superimposed on the interior where no interior cavity modes exist.

Additionally, observe in the display, the pressure decay associated with this predominately subsonic response at 88.5 Hz. The subsonic interior pressure decay between the cylindrical shell surface, a, and an interior radius, r, may be expressed in the helical spectrum as,[29]

$$P_n(r,k_z)=\frac{J_n(k_r r)}{J_n(k_r a)}P_n(a,k_z), \tag{1}$$

where

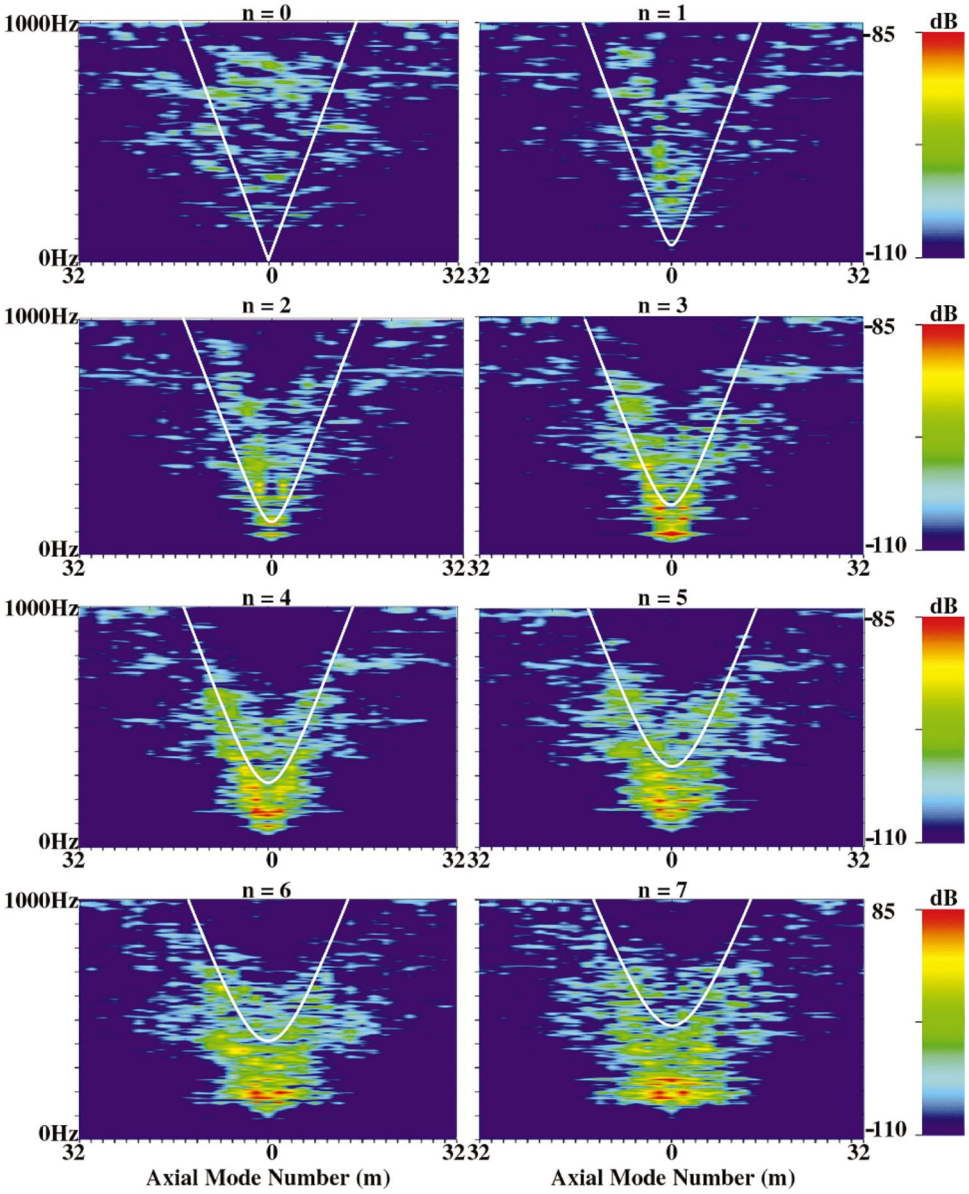


FIG. 9. Two-dimensional spatial Fourier decomposition of normal surface velocity response data due to a point excitation applied to the stiffener intersection (case 1). The white lines indicate the sonic transition from subsonic to supersonic motion for each mode set.

$$k_z = \frac{2\pi}{\lambda_z}, \quad k_r = \sqrt{k^2 - k_z^2}. \tag{2}$$

The pressure decay is a ratio of Bessel functions. For small arguments, as in the case of the $n=3$ $m=1$ mode, the Bessel function ratio can be approximated as

$$\frac{J_n(k_r r)}{J_n(k_r a)} \approx \left(\frac{r}{a}\right)^n, \tag{3}$$

indicating that pressure on a cylindrical interior will approximately follow a power law decay rate. Calculation of the Bessel function ratio is shown along with the experiment in Fig. 11 for the $n=3$ $m=1$ mode and for a sum dominant $m=1$ modes from $n=-5$ to 5 at 88.5 Hz. In the mode sum, the mode components have been scaled with experimental data to compensate for their relative amplitudes. The curves show that the $n=3$ structural mode dominates the interior response until about $r=45$ cm. The dominance of this subsonic mode in the pressure field is over a relatively long distance from the fuselage surface due its long structural wavelength. Higher order modes will decay faster as can be seen by examining the Bessel function ratio. Closer to the fuselage center ($r<45$ cm), lower order mode components, some of which are supersonic, dominate the response. The Bessel function ratio actually increases for $n=0$ toward the center of the fuselage and becomes dominant. This increase of the $n=0$ mode can be seen toward the center of the experimental circular pressure slice shown in Fig. 10. Also, close observation of the surface velocity dispersion curves of Fig. 9 reveals a faint hint of a supersonic response that can be seen in the $n=0$ and $n=1$ displays at 88.5 Hz. It is somewhat remarkable that these low level supersonic responses can dominate the interior response toward the center of the fuselage.

Interior rigid wall modes for this fuselage section are calculated in Houston and Herdic[30] using nomenclature where n is the number of full wavelengths around circumference, m is number of half wavelengths along the length, and s is the number node circles in the radial direction. The acoustic mode density, based on these calculations, is illus-

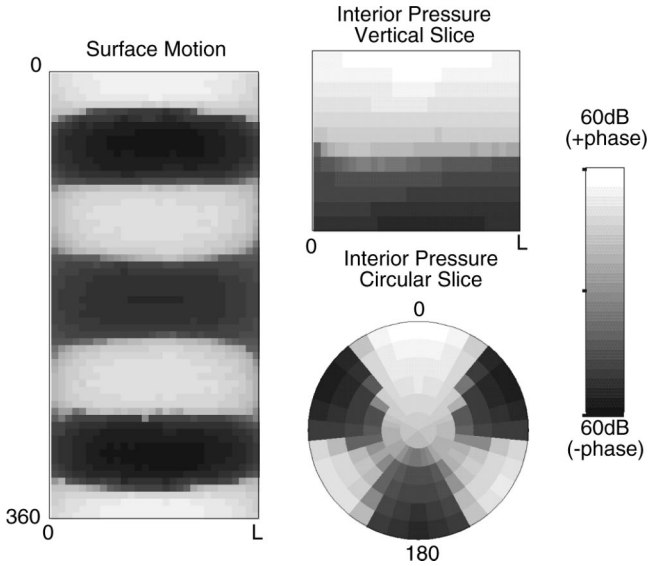


FIG. 10. Display surface motion [(m/s)/N] and interior pressure (Pa/N) of $m=1$, $n=3$ mode (88.5 Hz) for a point force applied to the fuselage intersection (case 1). Clear spatial matching of the interior response and the surface can be observed, demonstrating a structural mode forcing itself on the interior.

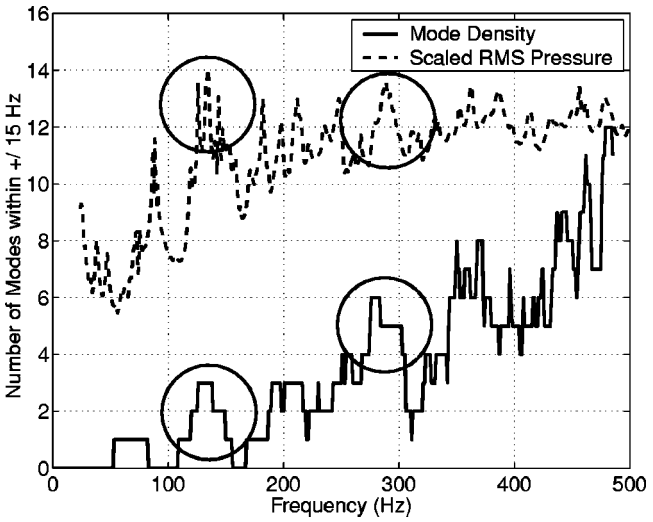


FIG. 12. Display illustrating the acoustic mode density for the fuselage section over a band from 0 to 500 Hz. The display indicates the number of acoustic modes within ± 15 Hz. Also shown is a scaled version of the rms interior response [case (1)] illustrating the correspondence between mode density and the interior pressure level at 130–150 Hz and around 275 Hz.

trated in Fig. 12. The display shows a moving sum of the number of modes within a ± 15 Hz range up to 500 Hz, along with a scaled version of the rms pressure curve under case (1) excitation. Observe the frequency shape dependence between the mode density and the rms interior response around $\sim 130-150$ Hz, and again where there is a mode cluster around 275 Hz. Recall from the structural response that individual panel motion occurs at these mid and higher frequencies. This local motion of the fuselage surface tends to act as local acoustic volume velocity sources, strongly exciting the interior acoustic modes. These volume velocity sources, combined with the higher acoustic mode densities, raise and maintain a high interior rms pressure at the mid to high frequencies with negligible evidence of the structure forcing the interior.

The rms spatial average of the interior pressure is shown in Fig. 13 for the three different excitations. Again, note that the response associated with the drive point excitations and external acoustic source have different units; it is only coincidental that the curves lie near the same magnitude level. Comparison of the interior acoustics under these three excitations again shows that the point force applied to the stiffener intersection (case 1) appears relatively strong throughout the band. Further, the case 1 response is a mixture of both structurally forced interior responses (at 88.5 Hz and 132.4) and interior cavity modes. The interior acoustic levels for the case 2 excitation are again very focused at low frequency, as was the case in its structural response due to the soft mount impedance and localized response within the

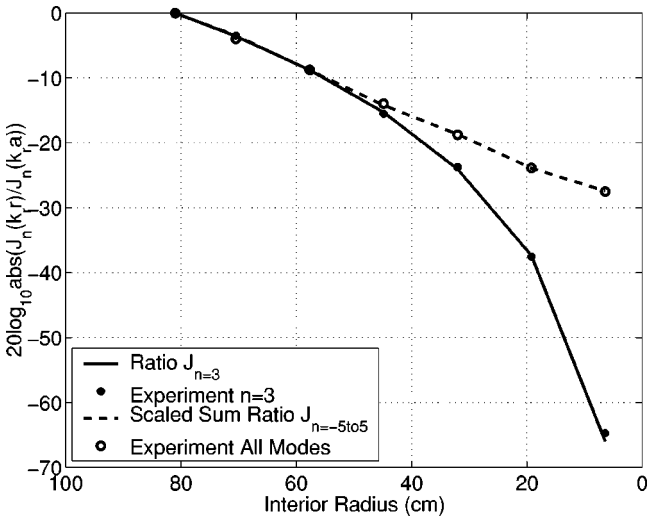


FIG. 11. Ratio of Bessel functions showing the interior pressure decay rate at 88.5 Hz for mode $n=3$, $m=1$ of a cylindrical shell. Also shown is the "experiment scaled" sum of Bessel function ratios for modes $n=-5$ to 5, $m=1$ that are dominant in the response. These calculated results are in good agreement with experiment.

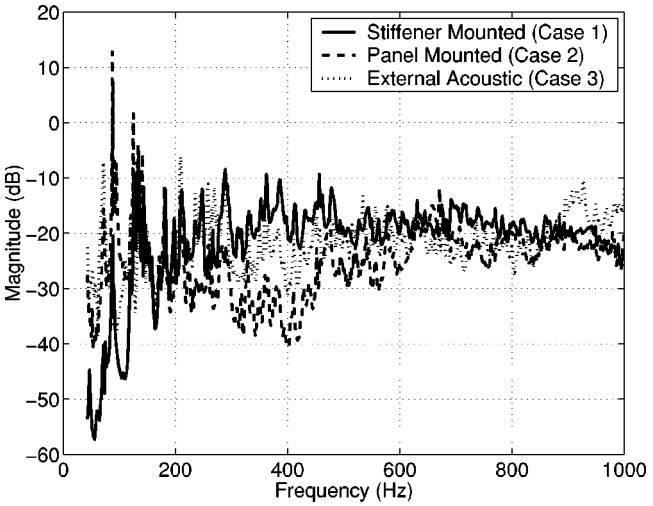


FIG. 13. rms interior pressure due to three excitation cases. A point force [(m/s)/N] applied at a rib/stringer stiffener intersection (case 1), a point force [(m/s)/N] applied at a panel (case 2), and external acoustic [(m/s)/Pa] excitation (case 3).

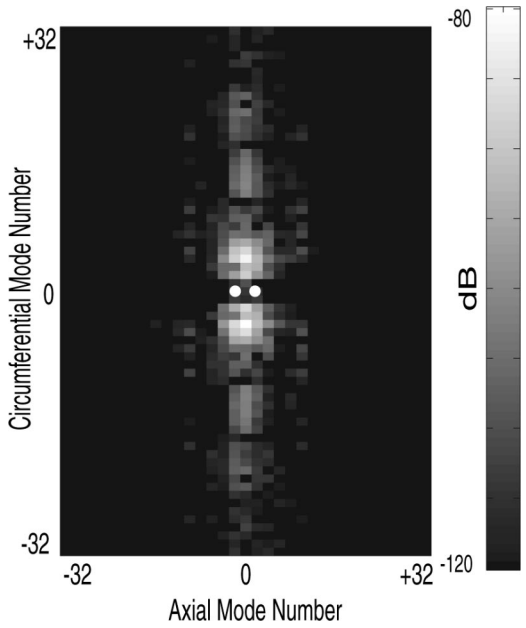


FIG. 14. Structural mode map at 134.2 Hz where the white dots indicate the strongest modes that can couple to the interior based on the spatial acoustic mode distribution (0,2,0) and frequency proximity criteria. Note that the vibration levels are very low at these modes, indicating that there is little interior noise generated due to the mode coupling mechanism.

driven panel. The interior response due to the external acoustic source (case 3) is completely dominated by the cavity modes, where the result of structurally forced modes on the interior is not observed. For example, no significant interior response can be found in the case 3 curve where the strong structural $n=3$ mode exists at 88.5 Hz frequency. The structural response to the acoustic source is simply weak and does not drive the interior.

Spatial coupling of structural modes to interior cavity modes is negligible in these results. In addition to spatial mode matching, the modes must also meet a frequency proximity requirement. The frequency proximity defined by Fahy[11] and Henry *et al.* [12] is rewritten here in terms of frequency bounds on the acoustic mode,

$$\frac{f_s(1-\eta_s/2)}{(1+\eta_a/2)} < f_a < \frac{f_s(1+\eta_s/2)}{(1-\eta_a/2)}, \tag{4}$$

where f_s and f_a are the mode frequencies of the fuselage structure and the interior acoustic cavity, respectively. η_s and η_a are the associated modal loss factors which have been determined experimentally by the half-power bandwidth method at single mode frequencies. Very few acoustic modes of this fuselage section fall within the frequency proximity criteria of the structural modes at low frequency. For example, there are no acoustic modes in frequency proximity to the strong $n=3$, $m=1$ mode at 88.5 Hz ($\eta_s=0.016, \eta_a=0.007$), and hence, no mode coupling. There is one acoustic mode ($n=0$, $m=2$, $s=0$) in frequency proximity to the structural response ($n=4$, $m=2$) at 134.2 Hz ($\eta_s=0.025, \eta_a=0.007$), but the strengths of the spatially matched structural modes are extremely weak in amplitude. The structural mode map for 134.2 Hz is shown in Fig. 14 where the white dots indicate the structural modes that meet the acoustic coupling criteria. The magnitude level of these structural modes is only ~2.0% of the maximum response at 134.2 Hz, and only ~0.2% of the response when compared to all structural modes below 500 Hz. A third example, the next strongest structural response at 209.4 Hz ($\eta_s=0.026, \eta_a=0.007$) again does not have any acoustic modes in frequency proximity. All of the other modes below 500 Hz have been analyzed in this fashion, and no modes show any greater coupling capability than the examples above.

IV. NUMERIC MODELING

A numerical database was generated by a finite element modeling approach that included the calculation of structural displacements and interior fluid pressures developed in response to the sidewall force. The calculation employed a finite-element/infinite-element, frequency domain formulation called SARA.[31] As with typical finite element calculations, the region is modeled with many small volume elements. These elements have matrix equations of motion derived by imposing conservation of energy. The matrices of each element are assembled to yield one global matrix equation. A detailed description of the model is beyond the scope of this paper. For that, the reader is referred to Allik *et al.*,[31] perhaps augmented for background by Zienkiewicz.[32]

With the SARA software, finite elements can model the structure, the interior fluid and the exterior acoustic nearfield of the fluid media. Infinite acoustic elements[33] are able to model the unbounded external fluid with a formulation that incorporates both farfield supersonic traveling waves and nearfield evanescent decaying waves. The excitation for these predictions corresponds to the experimental case 1, where a point force is applied to a rib/stringer stiffener intersection at one end of the fuselage.

A. Finite element model configuration

An axisymetric SARA two-dimensional (2D) model was considered early in the modeling process, but had significant limitations. This observation is somewhat obvious given the complexity of the structure. However, in comparing the axisymetric model with the experiment and full three-dimensional (3D) model, some important structural features of the fuselage were correlated to the response. For example, the strong $n=3$ mode was not observed in the 2D calculated result. The 2D model can only incorporate the fuselage skin and frames properly, but the stringers, window sections, and endcaps must be smeared over the area that they cover. When the window sections were added in the 3D model, the strong $n=3$ mode appeared as in the experiment. This exercise demonstrated the requirement for 3D modeling, and the significance between the strong $n=3$ mode response and the window section spacing at one-third of the circumference.

The development of the full SARA-3D model started out by using some basic estimates obtained from the exact solutions of simpler problems. For example, to determine the element sizes, the wavelengths traveling in the air and on the cylindrical shell must be estimated for the highest frequency of interest. As a starting point, the highest experimental frequency of 1000 Hz is used. The wavelength in air is 0.34 m; thus, a quadratic acoustic element should be no greater than one-fourth of this value. The dispersion relation for an infi-

nite cylindrical shell with smeared stiffeners is easy to compute[24] and reveals a wavelength of 0.31 m at 1000 Hz. Structural element sizes are chosen to accommodate the smaller of these two wavelengths. Finite element runs are then made with such a mesh and with finer meshes until convergence of the solution is achieved. For this model, a structural element size of 0.042 m was found to be adequate up to 250 Hz. The acoustic elements at the interface must be the same size as the sturctural elements so that they may mate and couple energy. This makes the acoustic elements smaller than required, but sufficient for the calculation. The other fundamental choice deals with the use of beam, shell or solid elements for the stiffeners and fuselage skin. Computation of the lowest resonance of a typical Cessna stiffener yields a frequency of ~485 Hz. Since this frequency is well within our band, beam elements are inadequate for representing the stiffeners. Beam elements are also inadequate to model the fuselage skin because of its two-dimensional nature. The choice of shell elements instead of solid elements is a simple one since the aluminum compressional wavelength is about 5 m at 1000 Hz, ~3000 5000 times the thickness of the stiffeners or fuselage skin.

After these initial estimates were made, preliminary experimental results revealed important physics that prompted the inclusion of extra complexity in the model. By employing this approach, complexity is gradually added to the model in an intelligent fashion until sufficient structural detail is included. This process leads to a very efficient model that does not contain any unnecessary model detail, yet can obtain accurate predictive results. It is also important to stress that the changes made to the model must be consistent with the vibro-acoustic physics of the fuselage in order to achieve good agreement at all frequencies. The process that was developed here for the fuselage section in this paper lends itself to a general model development procedure that can accurately predict the response of structures with moderate to high complexity.

The importance of experimental data and access to a physical model for accurate model development is now discussed in some simple examples.

To begin the process, detailed structural dimensions of the physical model must be obtained for the modeling effort to be successful. Initially, no drawings existed for the fuselage in this paper. Detailed dimensional measurements were required involving the overall size of the fuselage, stiffener size and location, the endcaps, etc. Ultrasonic measurements also revealed two thicknesses around the circumference. In the end, a detailed drawing set was generated that could be used by the modeler.

To check the finite element model, its total mass was compared to that of the physical model. An early model had a mass of 497 lb, and the fuselage structure measured 560 lb. The 63 lb of missing mass had a nontrivial effect on the response. It was determined that the missing mass was due to sealant and rivets not included on the circumferential portion of the fuselage model. By adding the additional mass to the model, the model's eigenfrequency associated with the $n=3$ mode shifted to within 1% of the experimental result, an 18% increase in the accuracy of the mode frequency prediction. In a very similar manner, the precise distribution of mass and stiffness of the shell endcaps that make up the end boundary conditions was also found to be significant in predicting these eigenfrequencies.

The structural damping determined through experimentation also played a vital role in the model development. Single mode measurements of the fuselage section showed loss levels significantly higher than the intrinsic material loss. These increased losses are due to phenomena such as the friction associated with the riveted construction and radiation damping losses to the fluid. It is necessary that the structural loss factor of the numeric model contains all losses associated with the structure, that is, inherent material damping, all riveted joint losses, etc. This value was obtained by taking the difference between the experimentally measured loss (structural and radiation loss) and the loss due to radiation damping alone. The level of radiation damping was estimated by examining the SARA-3D predicted response in the presence of fluid (air) and *in vacuo*, where radiation damping is not allowed. This derived structural loss factor was found to be ~0.16 in a frequency band up to 800 Hz with an exception in a small band between ~140 and 170 Hz where the value increased slightly to 0.20–0.26 range. The dominant nature of the 0.16 loss value afforded its use in the numerical predictions. Also note that it was very apparent in this analysis and previous work[24] that *in vacuo* models cannot be used to approximate the air-loaded case. Although the calculation time is significantly reduced by using an *in vacuo* model, the results are in dramatic error since radiation damping is not permitted.

The scope of the final SARA-3D model was very large, reflective of the level of detail required of both the fuselage structure as well as the interior and exterior fluid. The structural portion of the model, shown in Figure 15, involved an input deck that consisted of 4546 quadratic elements. Taking advantage of symmetry, only one fourth of the entire fuselage section was modeled. Shown in the zoomed view of Figure 15 is a typical fuselage bay (area between frames and stringers) with a 6×6 shell element grid. The circumferential frame segments in each bay consist of 2 shell elements for the web height and 1 shell element for the flange while the stringer height consisted of a single shell element. Some stiffeners were simplified in shape, but an equal moment of inertia was maintained. The window areas were modeled with Plexiglas material parameters, replacing the fuselage skin, frames and stringers in that area.

The structural endcap is stiffer than the fuselage skin thus requiring a lower mesh density. However, the edges of the endcap mesh and fuselage skin mesh must mate and be physically coupled. Both of these goals are obtained by creating an irregular grid on the endcap near the intersection, and then, opening up the elements to be slightly larger toward the center. These larger elements reduce the total number of nodes and shorten the finite element calculation.

The display in Figure 16 shows the mesh of acoustic elements. The interior acoustic elements toward the end of the fuselage are mated to the structural endcap elements, and then, the elements are replicated along the fuselage length at the axial structural nodes to form the acoustic mesh. Mating

of the interior acoustic elements and the elements of the structure (fuselage and the endcap) allows them to be connected and to couple vibro-acoustic energy. The frequencies within this study are below the coincidence frequency [11], and therefore, the acoustic wavelengths are larger than the structural wavelengths, making the acoustic element refinement more than adequate. The finite exterior acoustic elements are concentric layer also coupled to the structure. Each layer is one fourth of an acoustic wavelength; the layer thickness varies with frequency, and it is automatically adjusted by the SARA program during the calculation. A convergence study over the frequency range of interest found that one finite layer and one infinite layer are sufficient. Figure 16 shows only the first finite layer of exterior elements, but beyond that is a coupled layer of infinite elements that exists to extend the fluid indefinitely.

This acoustic element mesh contained ~9800 interior elements and ~5300 exterior elements. Combined with 4546 quadratic structural elements, the final model (structure and fluid) contained ~150,000 degrees-of-freedom overall. This level of structural refinement is consistent with similar efforts [25, 26], but included in this study are acoustic elements coupled to the structure.

Through this model development process, the predictive accuracy was found to be sensitive to a number of aspects related to the structure including nonaxisymmetry, mesh resolution, mass distribution, window material parameters, and structural damping. However, some approximations were made in this model that did not significantly affect the result. These approximations include use of a symmetry assumption where one fourth of the fuselage was modeled an equal moment of inertia approximation of stiffener cross sections, and treatment of the riveted areas as if the materials are completely coupled in molecular contact. However, the structural loss associated with these riveted areas was included in the model by using a loss value determined through an experimental procedure.

In starting the validation process, the drive point admittance was used to make comparisons between experiment and model. The drive point admittance is a quick look comparison since it requires only a single point numeric calculation as well as single point measurement. It is important to get the drive point response modeled correctly so that good predictive accuracy can be achieved throughout the structure.

B. Numeric results

Numeric predictions on the fuselage section interior and surface were performed using a point force excitation (experimental case 1). The calculations were made over a band from 30 to 250 Hz using 1 Hz resolution. The drive point impedance for the model and experiment is compared in Fig. 17. The comparison at the drive point shows generally good agreement, except perhaps in the 140–170 Hz range. Recall that this frequency band has been affected by the decision to use the 0.016 structural loss factor over all frequencies in the model calculation. In this narrow band, the damping determined through the experimental procedure is really higher in the range of ~0.020–0.026. The use of a frequency dependent loss factor would improve the accuracy of the calculation in this band, but has not been performed to date.

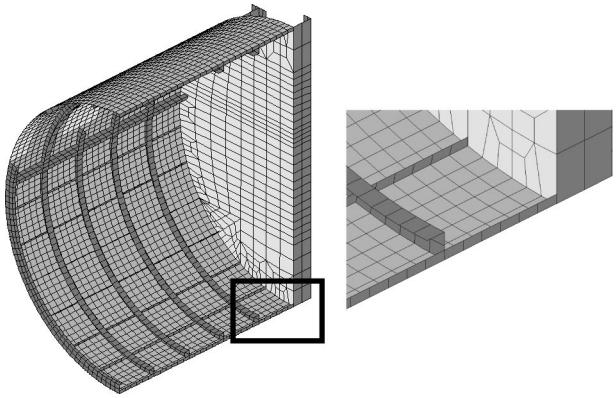


FIG. 15. The display of the left is an isometric view of the fuselage finite element structural mesh. Taking advantqge of symmetry, only on efourth of the actual fuselage is modeled. The light gray areas indicate the endcap and windows, the moderate dark areas are the fuselage sidewall, and stiffeners are shown with the darkest coloring. The box indicates a zoomed in area displayed on the right, showing detail of the frames and stringers of an individual bay.

A more detailed comparison of the surface velocity is shown in Fig. 18 where the displays illustrate the response as a function of circumferential and axial position. The experimental result is displayed on the left and the prediction on the right for four low frequency responses. The first comparison (upper left) is at the $n=3$, $m=1$ mode. The mode shape and frequency show excellent agreement, and the magnitude is only slightly higher in the experiment. In the bottom half of the fuselage ($90°-270°$), elevated levels in the predicted response are in a somewhat cross-hatched pattern. This pattern corresponds with the locations of the stiffeners. The experimental measurement points do not occur on the stiffeners, whereas the model mesh has points lying on all the ribs and stringers as well as within the bays. There is a slight hint of the cross-hatched pattern in the experiment since some measurements are made close to the stiffeners. The result of these mesh differences is again observed at the second comparison frequency at ~130 Hz (lower left). The response comparison is similar in mode shape and frequency, but

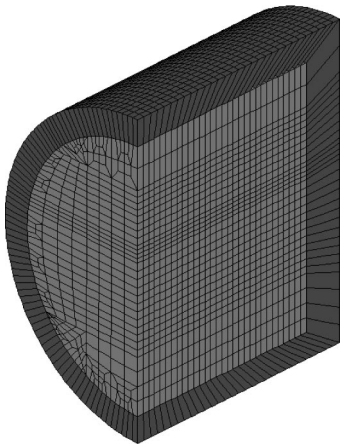


FIG. 16. Isometric view of the fuslage acousti element mesh. The fuselage structure is between the lighter color interior finite elements and darker exterior layer. Beyond the exterior finite layer is a layer of infinite elements (not shown) that extend the fluid indenfinitely.

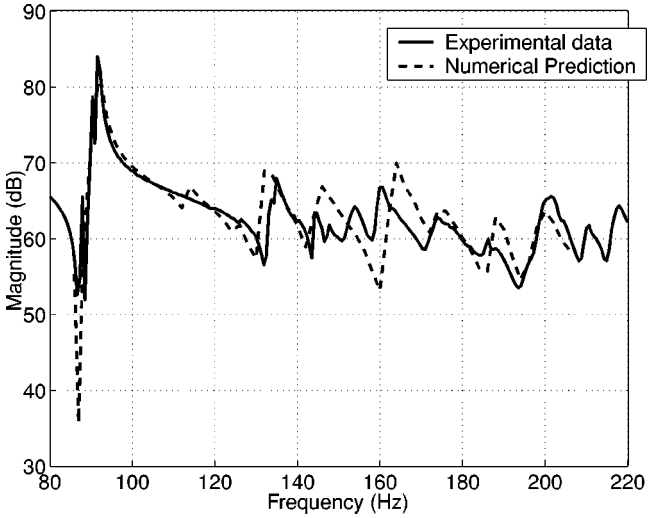


FIG. 17. Drive point impedance [N/(m/s)] comparison between the numerical prediction and the experimental result due to a point excitation applied to the stiffener intersection [case (1)].

again, a stronger cross-hatched pattern appears in the predicted response.

The other two frequency comparisons are for the strong responses at ~160 Hz (upper right) and ~208 Hz (lower right). In this mid-frequency regime, the model demonstrates that it has captured the response phenomenon associated with strong local resonances superimposed on a lower amplitude global mode. The local responses in the prediction appear to have a slightly different spatial positioning and level as compared to the experiment. The reason for these differences is unclear, but may be associated with the modeling of the window sections in the area. The oval windows have been modeled as square with the edges completely coupled or molecularly bonded to the aluminum fuselage shell. In actuality, the joining really involves additional framing, riveting, and sealant. Also, note that a modification of the damping to be more in line with the experimental result will improve the agreement at 160 Hz.

An interior pressure comparison between experiment (88.5 Hz) and numeric prediction (88 Hz) is shown in Fig. 19. The vertical and circular slices of the pressure field are

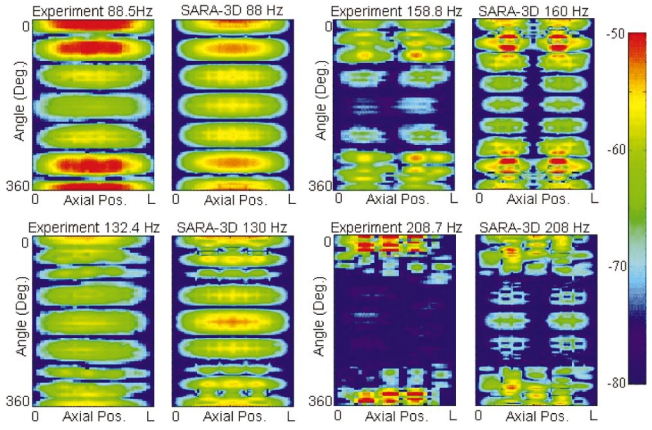


FIG. 18. Four detailed comparisons of surface velocity [(m/s)/N] magnitude distributions between the numerical prediction and the experimental response due to a point force excitation applied to the stiffener intersection [case (1)].

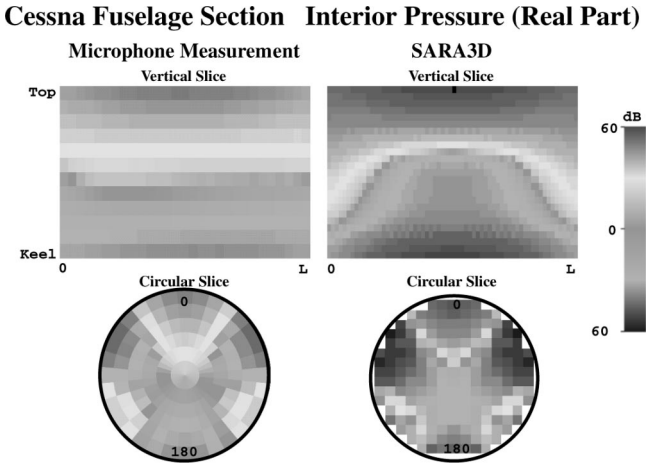


FIG. 19. Interior pressure comparisons (Pa/N) of the magnitude distribution between the numerical prediction and the experimental response at structural mode ($n=3,m=1$) due to a point force excitation applied to the stiffener intersection [case (1)].

taken through the center of the fuselage. The $n=3$ $m=1$ structural mode is shown here forcing the interior. Both the absolute pressure level and spatial distribution are in very good agreement. The only slight differences are at the phase transition near the ends of the interior space and that the numerical result shows little evidence of other mode components toward the center of the circular slice. The experiment has a clear $n=0$ component at the center as predicted by the Bessel function ratio in Eq. (1), but the numeric model exhibits a pure excitation of the $n=3$ mode. This may be due to a lower level of mode coupling between the $n=3$ mode and low order structural modes in the numeric model (reduced complexity) as compare to the physical model.

V. CONCLUSIONS

Several scanned laboratory measurements have been acquired, where normal surface velocity and interior pressure have been mapped out in a spatially dense manner under three separate excitations. The point force applied at the intersection of the rib and stringer (case 1) was studied in detail due to the rich vibro-acoustic physics that were excited over the band from 10 to 1000 Hz. This forcing function easily transmitted energy throughout the structure via coupling to the stiffener grid and excites the relevant vibro-acoustic mechanisms in a manner stronger than the other two cases.

The interior response at low frequency is forced by global structural modes driving the interior fluid. These strong responses are all subsonic, and the pressure decay rate away from the cylindrical fuselage surface follows a power law. The subsonic response, for example, at 88.5 Hz was found to dominate the interior until about $r=45$ cm. This distance is 36 cm from the fuselage surface and about 1/8 of an acoustic wavelength. Beyond 1/8 of the acoustic wavelength, lower order supersonic mode components take over, keeping pressure levels fairly high toward the fuselage center. At the mid and high frequencies, the normal acoustic modes dominate the interior response. The acoustic mode density begins to significantly increase in the 130–200 Hz range, and the

modes are easily excited by volume velocity sources created by local motion of the fuselage skin. It should be noted that this normal acoustic mode based mechanism will not be as dominant in a fuselage in which the cavity is broken up by partitions, trim, luggage compartments, seats, etc., and where acoustic absorption factors are larger. In a fully populated business aircraft, some normal modes will form in partitioned areas, but the dominant noise source is likely the low frequency mechanism associated with global structural modes and propulsion dynamics.

Clear frequency banding is observed in local one-dimensional line and ring normal surface velocity measurements. The objective of some work[34] is to exploit these banding effects to minimize noise transmission. However, when the disorder of the fuselage's periodicity is increased, as found in the two-dimensional wavenumber decompositions in this paper, the banding effects tend to diminish. Further, when additional disorder is added to the fuselage, such as in a fully populated aircraft interior, the banding effects will become even less evident. The low level of frequency banding associated with systems of high disorder, such as an aircraft fuselage, deems this noise control technique ineffective.

A symbiotic relationship exists between the numeric model and the physical model/experimental results. In the case of this fuselage section, the experimental response was better understood through the use of a numeric model as, for example, in identifying the significance of the circumference/3 spacing of the window sections and the strong $n=3$ modal response. The numeric model predictions, on the other hand, showed a high degree of accuracy that could not be achieved without a model development process fused to experimental results. This model has brought to bear a tool that aircraft manufacturers can confidently use to study the effect that structural modifications have on the interior sound field, and in general, a procedure that predicts the response of moderate to high complexity structures.

Since the vibro-acoustic physics of the fuselage section are well in hand, the progression of this work will involve implementation of interior noise quieting techniques. The focus will be to use a low mass treatment that combines performance from both passive and active control mechanisms. These control experiments will be conducted with a focus at low frequency where the disturbances are most intrusive and more difficult to treat because of the long wavelengths.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Richard Silcox, Structural Acoustic Branch Head, NASA, Langley and Robert Howes of Cessna Aircraft for their sponsorship, insight, and supply of the fuselage section used in this work.

[1] A. von Flotow and M. Mercadal, ''The measurement of noise and vibration transmitted into aircraft cabins,'' Sound Vib. Mag. **29**(10), 16–19 (1995).
[2] C.K. Barton, ''Cabin noise and weight reduction program for the Gulfstream G200,'' J. Acoust. Soc. Am. **112**, 2347 (2002).
[3] R. Howes, Cessna Aircraft (private communication).
[4] J. Wilby and F. Gloyna, ''Vibration measurements on airplane fuselage structures I. Turbulent boundary layer excitation,'' J. Sound Vib. **23**, 443–466 (1972).
[5] B. Magliozzi, D.B. Hanson, and R.K. Amiet, in *Propeller and Fan Noise*, Aeroacoustics of Flight Vehicles Vol. 1, edited by Harvey H. Hubbard (Acoustical Society of America, Woodbury, NY, 1995).
[6] L.D. Pope, C.M. Willis, and W.H. Mayes, ''Propeller aircraft interior noise model. II Scale-model and flight test,'' J. Sound Vib. **118**, 469–493 (1987).
[7] P. Goransson, ''Finite element calculations of the interior noise of the Saab 340 aircraft,'' Soc. Automoto Eng., SAE Paper 891081, 1989.
[8] L. Cheng, ''Fluid-structural coupling of a plate-ended cylindrical shell: Vibration and internal sound field,'' J. Sound Vib. **174**, 641–654 (1994).
[9] L. Pope, D. Rennisson, D. Willis, and W. Hayes, ''Development and validation of preliminary analytical models for aircraft interior noise prediction,'' J. Sound Vib. **82**, 541–575 (1982).
[10] S. Snyder and C. Hanson, ''The design of systems to control actively periodic sound transmission in to enclosed spaces. II. Mechanisms and Trends,'' J. Sound Vib. **170**, 451–472 (1994).
[11] F. Fahy, ''Vibration of containing structures by sound in the contained fluid,'' J. Sound Vib. **10**, 490–512 (1969).
[12] J. Henry and R. Clark, ''Noise transmission from a curved panel into a cylindrical enclosure: Analysis of structural acoustic coupling,'' J. Acoust. Soc. Am. **109**, 1456–1463 (2001).
[13] L. Cremer, M. Heckl, and E. Ungar, *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies*, 2nd ed. (Springer, Berlin, 1988).
[14] D. Mead, ''Vibration response of wave propagation in periodic structures,'' J. Eng. Ind. **93**, 783–792 (1971).
[15] C. Hodges, J. Powers, and J. Woodhouse, ''The frequency vibration of a ribbed cylinder. I. Theory,'' J. Sound Vib. **101**, 219–235 (1985).
[16] N. Ashcroft and D. Mermin, *Solid State Physics* (Brook Cole, Pacific Grove, CA, 1976).
[17] D. Photiadis, ''Acoustic behavior of near periodic elastic structures,'' Conference on Mechanical Vibration and Noise, 1995.
[18] L. Brillouin, *Wave Propagation in Periodic Structures*, 2nd ed. (Dover, New York, 1953).
[19] C. Hodges, ''Confinement of vibration by structural irregularity,'' J. Acoust. Soc. Am. **96**, 2291–2301 (1994).
[20] P. Anderson, ''Absence of diffusion in certain random lattices,'' Phys. Rev. **109**, 1492–1505 (1958).
[21] D. Photiadis, ''Anderson localization of one-dimensional wave propagation on a fluid loaded plate,'' J. Acoust. Soc. Am. **91**, 771–780 (1994).
[22] D. Photiadis and B. Houston, ''Anderson localization of vibration on a framed cylindrical shell,'' J. Acoust. Soc. Am. **106**, 1377–1392 (1999).
[23] D. Photiadis, E. Williams, and B. Houston, ''Wave-number space response of a near periodically ribbed shell,'' J. Acoust. Soc. Am. **101**, 877–886 (1997).
[24] B. Houston, M. Marcus, J. Bucaro, E. Williams, and D. Photiadis, ''The structural acoustics and active control of interior noise in a ribbed cylindrical shell,'' J. Acoust. Soc. Am. **99**, 3497–3513 (1996).
[25] R.D. Buehrle, G.A. Fleming, and R.S. Pappa, ''Finite element model development and validation for aircraft fuselage structures,'' 18th International Modal Analysis Conference, San Antonio, TX, 2000.
[26] F.W. Grosveld, J.I. Pritchard, R.D. Buehrle, and Richard Pappa, ''Finite element modeling of the NASA Langley Aluminum Testbed Cylinder,'' AIAA Paper 2002-2418, 2002.
[27] M.H. Marcus, B.H. Houston, and P.C. Herdic, ''Numerical modeling results of Cessna aircraft structures,'' Fourth AST Interior Noise Workshop, 2000.
[28] R.W. Hess, R.W. Herr, and W.H. Mayes, ''A study of the acoustic fatigue characteristics of some flat and curved aluminum panels exposed to random and discrete noise,'' NASA TN D-1, Aug 1959. (Also in Leissa, Vibration of Shells)
[29] E. G. Williams, *Fourier Acoustics* (Academic, Cambridge, UK, 1999).
[30] B. Houston, P. Herdic, M. Marcus, E. Williams, and J. Bucaro, ''High spatial density surface velocity and interior acoustic measurements associated with an aircraft fuselage section under point excitation,'' AIAA Paper 96-1765, 1996.
[31] H. Allik, R. Dees, S. Moore, and D. Pan, *SARA-3D User's Manual*, Version 94-1 (BBN Systems and Technologies, New London, CT, 1994).
[32] O.C. Zienkiewicz, *The Finite Element Method* (McGraw-Hill, New York, 1994).
[33] B. Bettess, *Infinite Elements* (Penshaw, Sunderland, UK, 1992).
[34] D. Mead and N. Bardell, ''Free vibration of a thin cylindrical shell with discrete axial stiffeners,'' J. Sound Vib. **111**, **499–520 (1986).**

Sound transmission across duct constrictions with and without tapered sections

C. K. Lau and S. K. Tang[a)]
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

(Received 29 March 2004; revised 29 March 2005; accepted 5 April 2005)

The sound power transmission loss across duct constrictions with linearly tapered sections is studied with the finite element method. Results show that the acoustic energy distributions of transmitted waves at high frequency depend critically on the exit configuration of the constriction. The corresponding strengths of these waves are very much affected by the entrance setup of the constriction. The difference between inlet and outlet impedance of a constriction leads to weaker resonant sound transmission. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1921549]

PACS numbers: 43.50.Gf, 43.20.Mv [DKW] Pages: 3679–3685

I. INTRODUCTION

Ventilation and air conditioning systems are major sources of noise in modern high-rise commercial buildings. Their noises propagate into the interior of buildings through the ventilation ductwork. However owing to practical reasons, these ductwork systems usually involve many ducted elements, such as constriction, expansion,[1] bending,[2] and tee-junction.[3] The prediction of sound transmission characteristics across these ducted elements is very important and thus, the topic has attracted the attention of many researchers in the past few decades. For instance, Miles[4,5] and Baumeister *et al.*[6] investigated the plane wave propagation across a variable area duct. Silcox and Lester[7] examined sound wave transmission across a conical constriction with a wall slope of 3.6°. Selamet and Easwaran[8] studied the plane wave propagation across a venturi tube of slow area variation in the presence of a mean flow using the stepwise expansion and contraction duct approximation. Cho and Ingard[9] discussed the propagation of higher modes across a constriction of slow and smooth area variation analytically while Hudde[10] examined similar propagation effects inside divergent and convergent sections again by the stepped duct approximation.

A recent study by the authors[11] revealed that the propagation characteristics of the acoustic modes across smooth convergent and divergent sections in an infinitely long duct are very different. However, the corresponding information for sound propagation across duct constrictions and the associated acoustic mode interactions have, to the knowledge of the authors, not been discussed in detail in existing literature unless drastic simplifications of the nonuniform section configurations are made (for instance, Refs. 9 and 10).

Tapering of the inlet and outlet of a constriction is recommended by building services engineers in order to reduce the fluid pressure loss along the ductwork and the level of regeneration noise. This also results in smoother acoustic impedance change than the abrupt constriction and thus diminishes the sound power transmission loss (TL). Since these effects are contradictory, the balance between them must be carefully optimized. However, it is currently uncertain on how the constriction geometry will affect the sound power transmission, especially when nonplanar waves exist in the duct. In addition, constrictions are found inside conventional dissipative silencers with porous materials. It is then essential to understand the sound wave propagation inside rigid wall constrictions before the corresponding effects of the porous materials can be fully understood.

In the present study, sound transmission across a constriction in an infinite duct is investigated by using the finite element method implemented by the software MATLAB. For simplicity, only linearly tapered convergent and divergent sections are studied. It is hoped that the present study will provide a deeper understanding on the sound transmission across constrictions inside ductwork and useful information for improved ductwork design in the future.

II. COMPUTATIONAL DOMAIN AND BOUNDARY CONDITIONS

Figure 1 shows the schematic diagram of the flow constriction adopted in the present study. The origin of the coordinate system is set at the center of the constriction section. For practical reasons, only two-dimensional linearly tapered convergent and divergent sections are considered. The geometry and acoustic excitation are symmetrical about the longitudinal duct axis and thus no asymmetric duct mode can be excited. The tapering angles of the convergent and divergent sections are denoted by ϕ_c and ϕ_d, respectively. The widths of the main duct and the flow constriction are denoted by d and w, respectively. The lengths of the flow constriction, convergent and divergent linearly tapered sections are denoted by L, $L_c[=(d-w)/2\tan\phi_c]$ and $L_d[=(d-w)/2\tan\phi_d]$, respectively. All the length scales are normalized by the width of the main duct d, and the frequency range investigated is up to $kd=9$, where k is the wave number of the sound involved.

The target here is to solve the following inhomogeneous wave equation with unit plane wave excitation:

$$\nabla^2 p + k^2 p = q, \tag{1}$$

a)Author to whom correspondence should be addressed; electronic mail: besktang@polyu.edu.hk

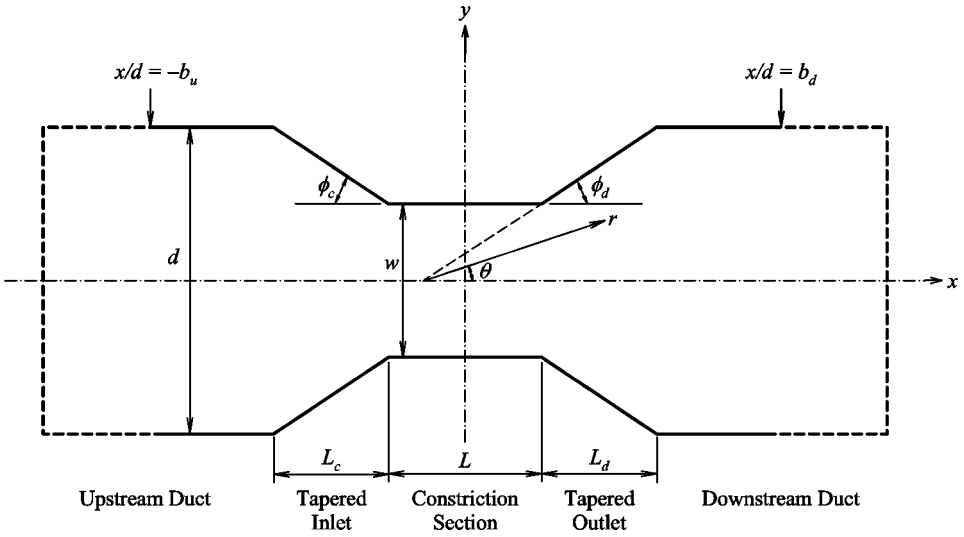


FIG. 1. Schematic diagram of the computational model.

where p and q are the acoustic pressure and sound source strength, respectively, with rigid wall boundaries so that the normal pressure gradients there vanish. For $kd<2\pi$, all the waves in the two straight duct sections are nominally planar. In this case, $q=0$. For a unit-strength acoustic plane wave incident at $x=-b_u$, the boundary conditions at the two ends of the computation domain are

$$\frac{\partial p}{\partial n}\pm ikp=\begin{cases}2ik, & x=-b_u\\ 0, & x=b_d\end{cases}, \tag{2}$$

where $i=\sqrt{-1}$, n is the outward normal direction of the boundary, b_u and b_d are the computational domain boundary locations (Fig. 1).

When $kd>2\pi$, higher modes are present in the upstream and downstream duct sections. The numerical treatment of Tang and Lau[11] is adopted to damp down the higher modes before they reach the end boundaries (dashed lines in Fig. 1) so as to minimize the numerical reflections there. The boundary conditions of the absorptive side walls are

$$\frac{\partial p}{\partial n}\pm ik\gamma p=0, \tag{3}$$

where γ is an artificial absorption coefficient which is set at $0.01(x+b_u)^2$ and $0.01(x-b_d)^2$ for upstream and downstream absorptive endings, respectively, as in Tang and Lau.[11] In this case, the unit plane wave excitation is[11]

$$q=2ik\delta(x+b_u), \tag{4}$$

where δ is delta function. Boundary conditions depicted in Eq. (2) apply at the extreme ends of the absorptive endings.

The TL for the plane wave in the present study follows the definition in standard textbook (for instance, Ref. 1). The TL for a higher duct mode is defined as the ratio of the higher mode energy to that of the incident plane wave. The total power transmission loss refers to the ratio of the total energy (plane wave plus higher duct modes) transmitted across the flow constriction to that of the incident plane wave.

All finite element computations are carried out using the partial differential equation (PDE) solver and the mesh generation facilities of the software MATLAB.[12] For each geometry of the constriction, triangular mesh is generated in the computational domain and the PDE (with the determined boundary conditions) is discretized to form a linear system $Ku_h=F$, where the matrices K and F depend on the coefficients of the PDE, and u_h is the unknown vector and contains the values of the approximate solution at the mesh points.

Meshes used in the present study are generated using the Delaunay triangulation algorithm[13] and are refined by the "regular refinement" scheme, where all the triangular mesh elements are divided into four triangles.[12] Figure 2 illustrates the computational mesh adopted for the case of $\phi_c=\phi_d$

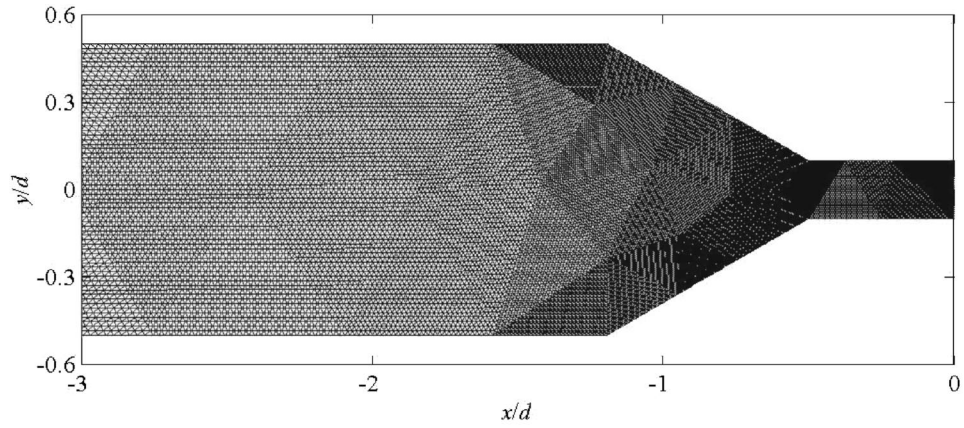


FIG. 2. Computational mesh for $\phi_c=\phi_d=\pi/6$ at $5<kd<9$.

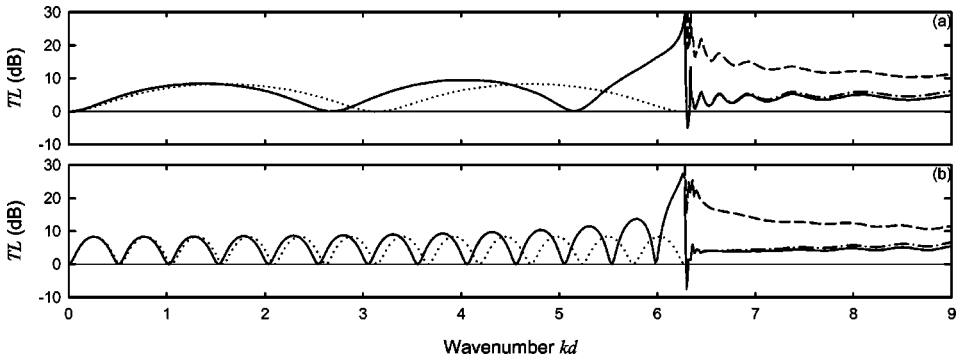


FIG. 3. Sound power transmission loss for constrictions with abrupt contraction and expansion. (a) $L/d=1$, $w/d=0.2$; (b) $L/d=6$, $w/d=0.2$. Plane wave theory (···); total power loss (—); (– – –) plane wave mode power transmission loss for $kd>2\pi$; (–·–) first symmetric higher mode power transmission loss for $kd>2\pi$.

$=\pi/6$ (only those within region $-3\leqslant x/d\leqslant 0$ are presented). The corresponding triangle quality m of each triangle is within 0.7 to ~1 which is satisfactory ($m>0.6$ is required[12]). At $kd\sim 9$, the corresponding minimum number of node points $n_{\min}$ per one wavelength of the sound is about 20.

The triangular mesh sizes and details vary with the geometries of the tapered sections and thus are not fixed throughout the investigation. However throughout the computations, m is kept greater than 0.6 and $n_{\min}$ not less than 20. Also, it has been confirmed that further refinement of the meshes will not give noticeable differences in the results.

III. RESULTS AND DISCUSSIONS

The present study investigates the sound wave propagation below the second symmetric cut-on frequency of the duct section ($kd<9$). The TL is estimated from the computed data at $x/d\sim 9$ where all the evanescent waves have been damped out completely.

Owing to the symmetrical excitation and the geometry of the constriction, only symmetrical mode will be excited. For $kd<9$, only the first symmetric mode inside the duct section will be excited while only plane wave propagation is observed inside the constriction section. Inside the tapered section, the corresponding symmetrical mode pattern is[11]

$$[A_\alpha H_\alpha^1(kr)+B_\alpha H_\alpha^2(kr)]\cos(\alpha\theta), \tag{5}$$

where r and θ are the radial position and the angular position, respectively (Fig. 1), $\alpha=j\pi/\phi$, where j is an integer greater than unity, and A_α and B_α are complex constants. H_α^1 and H_α^2 are the first and second kind Hankel functions of the order α, respectively. θ is bounded between $\pm\phi_c$ or $\pm\phi_d$.

For the convergent inlet, the sound wave begins to propagate in the form of the jth symmetric angular mode if $kd\geqslant 2(j\pi/\phi)\sin\phi$ and can propagate across such section only if

$$kd\geqslant \frac{2j\pi}{\phi}\frac{d}{w}\sin\phi.$$

Sound wave can propagate into the divergent outlet in form of the jth symmetric angular mode if kd satisfies the second requirement of propagation for the convergent inlet.

A. Constrictions with abrupt contraction and expansion

Figures 3(a) and 3(b) illustrate the effects of L/d on the TL across the constriction with abrupt contraction and expansion ($\phi_c=\phi_d=\pi/2$) at $w/d=0.2$. The TLs predicted by the plane wave theory[1] are also given for comparison. The computed TL is comparable to those obtained using the plane wave assumption for all L/d investigated when $kd<2\pi$. However, resonant transmission occurs at lower frequency than the plane wave theory prediction because of the effective length corrections at the two ends of the constriction section due to the radiation impedance there.[1,3]

At $kd\sim 2\pi$ (first symmetric duct mode cut-on frequency), the TL increases; a similar observation was found in the concentric expansion chambers of Selamet and Radvaich.[14] The higher TL near $kd=2\pi$ is due to the strong reflection back into the upstream duct section.

For $kd>2\pi$, the first symmetric higher duct mode is excited. It propagates along the duct sections together with the plane wave. The resulting sound amplification at kd just higher than 2π is due to the strong excitation of the first symmetric higher duct mode. An increase in the length of the constriction results in smaller TL fluctuations for $kd>2\pi$ [Fig. 3(b)]. In all cases with $w/d=0.2$, the higher mode dominates when $kd>2\pi$. The increase in w/d at a fixed L/d reduces the TL across the constriction as expected due to better impedance matching between the duct and the constriction sections. Thus, the associated results are not presented.

Figure 4 shows the sound wave interactions within the proximity of the constriction with abrupt contraction and expansion for $L/d=1$, $w/d=0.2$. At low frequency, the sound waves have circular wave fronts near the inlet and exit of the

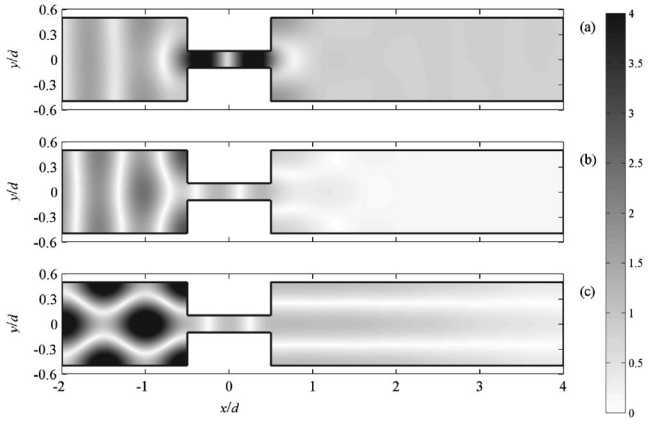


FIG. 4. Sound field patterns around constrictions with abrupt contraction and expansion. (a) $kd=5$; (b) $kd=6$; (c) $kd=6.3$. $w/d=0.2$, $L/d=1$.

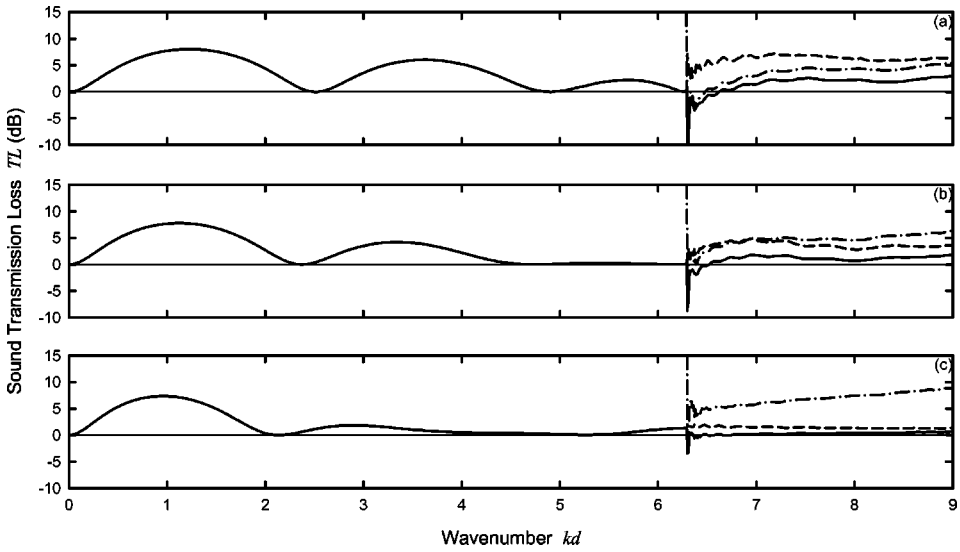


FIG. 5. Sound transmission across constrictions with tapered inlet and outlet. (a) $\phi_c=\phi_d=\pi/3$; (b) $\phi_c=\phi_d=\pi/4$; (c) $\phi_c=\phi_d=\pi/6$. $w/d=0.2$, $L/d=1$. Total power loss (—); plane wave mode power transmission coefficient for $kd>2\pi$ (– – –); first symmetric higher mode power transmission coefficient for $kd>2\pi$ (–·–).

constriction [Fig. 4(a)]. These wave fronts then spread out to fill the duct cross section with uniform energy density. At kd just below 2π, the first symmetric higher mode of the duct section is found at the exit plane of the constriction [Fig. 4(b)]. This mode decays rapidly as it propagates into the downstream duct section as expected. Strong acoustic reflection back into the upstream duct occurs resulting in the rapid increase of TL as kd approaches 2π [Fig. 4(b)].

The strong higher mode created by the resonance at $kd>2\pi$ carries the majority of energy into the downstream duct [Fig. 4(c)], resulting in sound amplification (Fig. 3). One should also note that the case of $kd\sim 2\pi$ coincides with one of the resonant transmission cases across the constriction. The sound field patterns at higher kd are similar to that shown in Fig. 4(c), except that more nodal and antinodal planes are observed because of reduced wavelength, and thus are not presented. Away from the resonance frequency, energy is gradually redistributed back to the plane wave mode, causing a reduction in the TL of the plane wave mode shown in Fig. 3 at increased kd.

The phenomena observed in Figs. 3 and 4 are also found in other values of L/d once w/d is fixed (not presented here). Therefore without loss of generality, the foregoing discussions concerning tapered sections are done with $L/d=1$ and $w/d=0.2$, while L_c/d and L_d/d vary with the angles of tapering ϕ_c and ϕ_d.

B. Constrictions with tapered inlet and outlet

With $L/d=1$ and $w/d=0.2$, the effects of ϕ_c ($=\phi_d$ in this section) on the sound transmission across the constriction are shown in Fig. 5. One would expect that a smaller ϕ_c will provide a better acoustic impedance matching between the constriction and the duct sections, resulting in a lower TL than that obtained with abrupt expansion and contraction for $kd<2\pi$. The improvement in such matching increases with frequency and thus the TL dome magnitudes decrease at reduced ϕ_c. At very low frequency where the wavelength of the sound is much longer than the lengths of the tapered sections, the sound wave treats these sections as abrupt contractions and expansions. The TL is thus approximately equal to the plane wave theory prediction. Increasing frequency (reducing the wavelength) allows better development of wave propagation inside the constriction section, resulting in a gradual drop of the TL as kd approaches 2π for a fixed constriction geometry. This situation is enhanced by reducing ϕ_c and thus the magnitudes of the TL domes decrease more rapidly with frequency in such cases. The TL therefore does not rise up when kd approaches 2π as in Fig. 3.

The first symmetric angular mode is excited when $kd=1.65\pi(\sim 5.18)$, $1.8\pi(\sim 5.65)$, and $1.91\pi(\sim 6.00)$ for ϕ_c ($=\phi_d$) equals $\pi/3$, $\pi/4$, and $\pi/6$, respectively. This higher angular mode matches better with the first symmetric higher duct mode and thus the latter is more efficiently excited than in the abrupt expansion case. A higher possibility of sound amplification at $kd\sim 2\pi$ is thus expected. However, the smoother development of the waves inside the tapered sections, especially those within the outlet, provides more uniform acoustic excitation to the duct modes. Figure 5 also illustrates that the contribution of the higher duct mode is reduced while that of the plane wave mode is enhanced as the angle of tapering decreases.

Within the ranges of frequency and angle of tapering in the present study, sound wave cannot propagate through the constriction in the form of the higher angular mode as kd is always below

$$\frac{2j\pi}{\phi}\frac{d}{w}\sin\phi.$$

At $kd\sim 6$, ($>1.65\pi$), the first symmetric angular mode can be observed only at the junction between the duct and the convergent inlet [Fig. 6(a)]. The wave patterns well inside the tapered sections remain very circular. Though the higher duct mode decays as it propagates in the downstream duct section, it penetrates longer into this duct section than in the cases of abrupt expansion [cf. Fig. 4(a)]. A reduction of the angle of tapering lengthens this penetration (not shown here).

At $kd>2\pi$, the first symmetric higher mode propagates in the duct sections. The good matching between the higher modes in the tapered sections and the duct sections is manifested in Fig. 6(b). The strong higher duct mode due to reso-

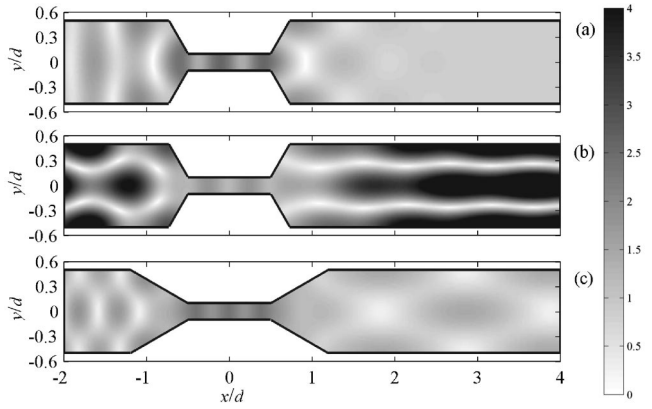


FIG. 6. Sound fields near constrictions with tapered inlet and outlet. (a) $\phi_c=\phi_d=\pi/3$, $kd=6$; (b) $\phi_c=\phi_d=\pi/3$, $kd=6.3$; (c) $\phi_c=\phi_d=\pi/6$, $kd=8$. $w/d=0.2$, $L/d=1$.

nance is also revealed. As the angle of tapering is reduced, the sound pressure magnitude becomes more uniform across the cross section of the downstream duct even at kd far higher than 2π. The diamond shape pattern of the sound field in Fig. 6(c) with its maximum and minimum values on the duct centerline further demonstrates the more dominant plane wave mode at reduced angle of tapering shown in Fig. 5.

C. Constrictions with one end tapered

The constriction with tapered inlet and abrupt expansion ($\phi_d=\pi/2$) produces higher TL (Fig. 7) compared with the one with both tapered inlet and outlet (Fig. 5), but gives lower TL compared with the constriction with abrupt contraction and expansion [Fig. 3(a)]. The maximum TL below the cut-on frequency of the first symmetric higher duct mode is again lowered at smaller ϕ_c. However, it is found that resonant transmission across the constriction becomes less likely here even at a frequency far lower than the first symmetric higher duct mode cut-on frequency as the angle of tapering ϕ_c is reduced. No vanishing TL can be found at $kd>1$ for $\phi_c=\pi/6$ [Fig. 7(b)]. The difference in the inlet and outlet impedance of the constriction creates opposite propagating plane waves of significant different magnitudes such that complete constructive interference cannot take place at the two ends of the constriction region. This is similar to the case of standing wave between two walls of different surface sound absorption.[15]

It is also found from Fig. 7 that the strength of the higher duct mode downstream of the constriction is strong compared to those illustrated in the previous section. The tapered inlet enhances the wave propagation into the constriction and eventually leads to stronger excitation of the higher duct mode at the exit plane of the constriction due to the singularity of the abrupt expansion. Both the strengths of this higher duct mode and the plane wave increase as the angle of tapering decreases as expected, but the former carries a higher percentage of acoustical energy and this percentage increases slightly when the rate of convergence (ϕ_c) is reduced. This is not the case for constrictions with tapered inlet and outlet (Fig. 5). Besides, the smoother convergence leads to the amplification of sound at high frequency [Fig. 7(b)]. The frequency range of such sound amplification is also widened at reduced ϕ_c (not shown here).

For the constrictions with abrupt contraction and tapered outlet ($\phi_c=\pi/2$), the TLs for $kd<2\pi$ are the same as those for the constrictions with tapered inlet and abrupt expansion (Fig. 8). It can also be observed that the TL is always positive here over the frequency range considered in the present study, except at frequencies close to the resonance of the first symmetric higher duct mode. However, the magnitudes and the cycles of the TL oscillations after this resonance are significantly higher than those in Fig. 7. The frequency of the TL oscillation does not appear to depend on ϕ_d. However, the tapered outlet favors energy transfer into the plane wave in the downstream duct section, while the abrupt expansion results in higher energy transfer into the higher duct mode (Figs. 3 and 7).

The relative importance of the higher duct mode at $kd>2\pi$ decreases at reduced ϕ_d. At $\phi_d=\pi/3$ [Fig. 8(a)], both the higher duct mode and the plane wave inside the downstream duct are of comparable magnitudes. As ϕ_d decreases, the more uniform excitation at the exit plane of the tapered outlet appears to favor the propagation of plane wave in the downstream duct [Fig. 8(b)]. This is opposite to what happens in the constrictions with tapered inlet and abrupt expansion, where the singularity at the abrupt expansion outlet favors the excitation of higher duct modes. Under this constriction configuration, the TL is always positive, except when kd is very close to 2π.

The characteristics of the corresponding sound fields for the constrictions with one end tapered are very similar to those presented earlier in this section and thus are not pre-

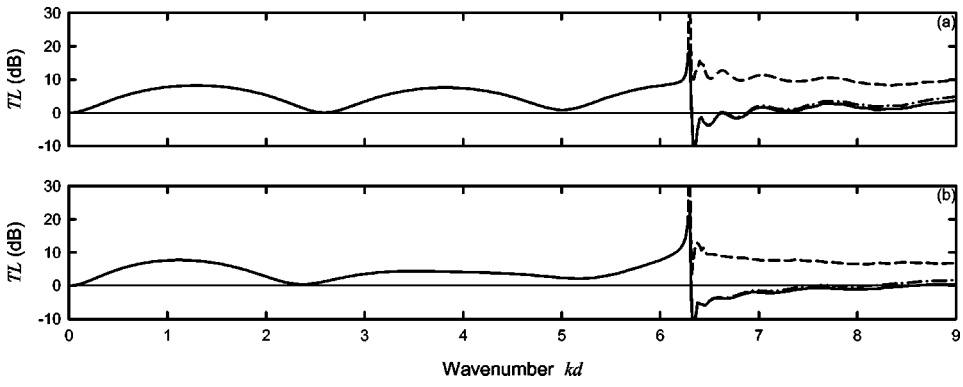


FIG. 7. Sound power transmission loss for constrictions with tapered inlet and abrupt expansion. (a) $\phi_c=\pi/3$, $L_c=0.23$, (b) $\phi_c=\pi/6$, $L_c=0.69$. $\phi_d=\pi/2$, $w/d=0.2$, $L/d=1$. Legends: same as those in Fig. 5.

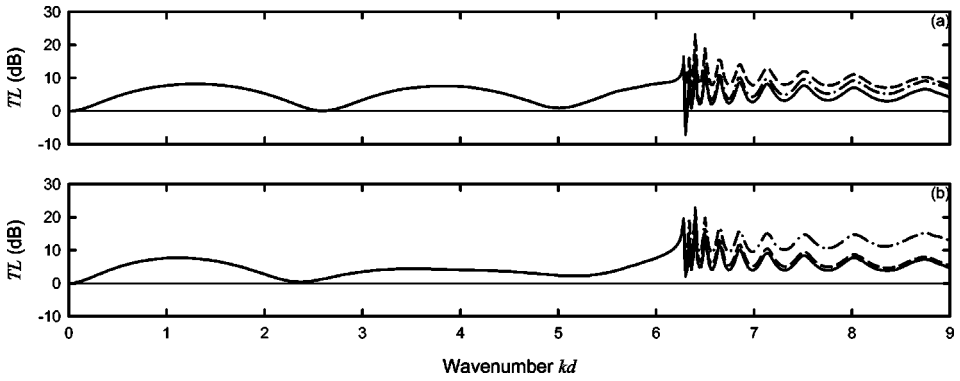


FIG. 8. Sound power transmission loss across constrictions with abrupt contraction and tapered outlet. (a) $\phi_d = \pi/3$, $L_d = 0.23$; (b) $\phi_d = \pi/6$, $L_d = 0.69$. $\phi_c = \pi/2$, $w/d = 0.2$, $L/d = 1$. Legends: same as those in Fig. 5.

sented. The main difference is only the relative magnitudes of the acoustic modes.

IV. PRACTICAL DESIGN CONSIDERATIONS

The tapering of the inlet and outlet of a constriction can reduce the fluid pressure loss and the level of aerodynamic noise generated, but at the same time this reduces the TL. A careful consideration of these effects is essential for practical design.

The pressure loss across the constriction, Δp, in duct can be estimated using

$$\Delta p = [(d/w)^2 C_i + C_o]\rho V^2/2, \tag{6}$$

where ρ is the density of the air, V the mean flow velocity in the main duct, and C_i and C_o the pressure loss coefficients of the inlet and outlet, respectively. According to the pressure-loss coefficients tabulated in ASHRAE,[16] the pressure loss across the convergent inlet is lower than that across the divergent outlet for the same tapering angle. Such difference increases with decreasing w/d. The variation of C_i is very small for $\phi_c < \pi/4$. C_o peaks at around $\phi_d = \pi/4$ and then decreases slowly as ϕ_d increases. Considering the results shown in Figs. 5, 7, and 8, the constrictions with abrupt contraction and tapered outlet appear to be able to give a compromise between the pressure loss and the TL as long as the aeroacoustics is ignored. The abrupt contraction causes smaller pressure loss than the abrupt expansion for $w/d \leqslant 0.5$, but can maintain a more stable and higher TL over the frequency range concerned. Though the abrupt contraction does result in a higher pressure loss than the tapered inlet, the broadband gain in the TL compensates such disadvantage. The constrictions with tapered inlet and abrupt expansion are likely to produce a higher pressure loss without improving the TL. The constrictions with abrupt contraction and expansion give the worst pressure loss without significant improvement in the TL and thus are not desirable.

The corresponding aeroacoustics is complicated. The power of the aerodynamic noise and its frequency depend substantially on the velocity of and the turbulence level in the main flow[17] as well as the geometry of the constriction. This part of the study is very involved and needed to be treated specifically in a separated investigation.

V. CONCLUSIONS

In the present study, the sound power transmission losses through flow constrictions with and without a linearly tapered inlet/outlet in an infinitely long duct are studied with the finite element method. The associated sound energy distributions between various acoustic modes are also discussed.

For the constriction with abrupt contraction and expansion, the computed sound power transmission loss agrees with the plane wave theory prediction at low frequencies. At frequency higher than the first symmetric duct mode cut-on frequency, oscillations of the overall sound power transmission losses are observed. The strength of the plane wave relative to that of the higher duct mode increases at increased frequency, but the latter remains the dominant acoustic mode up to the frequency limit of the present study.

The results in the present study also show that the sound power transmission loss across a flow constriction in an infinitely long duct before the cut-on of the first symmetric higher duct mode is lowered by a smoother constriction transition. Above the cut-on frequency of this higher mode, amplification of sound is likely when the inlet of the constriction is tapered. Such amplification is more significant at reduced tapering angle. Also, the relative importance between the plane wave and the higher duct mode inside the downstream duct depends only on the constriction outlet. For a smoother outlet, the higher mode is more dominant than the plane wave, while the opposite is observed if the constriction consists of an abrupt expansion.

ACKNOWLEDGMENT

This study is supported by a grant from the Research Grant Council, The Hong Kong Special Administration Region, People's Republic of China (Project No. PolyU5030/00E).

[1] L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics* (Wiley, New York, 1962).
[2] A. Cummings, "Sound transmission in curved duct bends," J. Sound Vib. **35**, 451–477 (1974).
[3] S. K. Tang, "Sound transmission characteristics of Tee-junctions and the associated length corrections," J. Acoust. Soc. Am. **115**, 218–227 (2004).
[4] J. H. Miles, "Acoustic transmission matrix of a variable area duct or nozzle carrying a compressible subsonic flow," J. Acoust. Soc. Am. **69**, 1577–1586 (1981).
[5] J. H. Miles, "Verification of a one-dimensional analysis of sound propa-

gation in a variable area duct without flow,'' J. Acoust. Soc. Am. **72**, 621–624 (1982).
[6] K. J. Baumeister, W. Eversman, R. J. Astley, and J. W. White, ''Acoustics in variable area duct: Finite element and finite difference comparisons to experiment,'' AIAA J. **21**, 193–199 (1983).
[7] R. J. Silcox and H. C. Lester, ''Sound propagation through a variable area duct: Experiment and theory,'' AIAA J. **21**, 1377–1384 (1982).
[8] A. Selamet and V. Easwaran, ''Wave propagation and attenuation in Herschel-Venturi tubes,'' J. Acoust. Soc. Am. **101**, 936–942 (1997).
[9] Y. C. Cho and K. U. Ingard, ''Mode propagation in nonuniform circular ducts with potential flow,'' AIAA J. **21**, 970–977 (1983).
[10] H. Hudde, ''Acoustical higher-order mode scattering matrix of circular nonuniform lossy tubes without flow,'' J. Acoust. Soc. Am. **85**, 2316–2330 (1989).
[11] S. K. Tang and C. K. Lau, ''Sound transmission across a smooth noniuniform section in an infinitely long duct,'' J. Acoust. Soc. Am. **112**, 2602–2611 (2002).
[12] L. Langemyr, A. Nordmark, M. Ringh, A. Ruhe, J. Oppelstrup, and M. Dorobantu, *Partial Differential Equation Toolbox User Guide* (The MathWorks Inc., MA, 1996. The 2004 version is available on-line at http::ww.mathworks.com).
[13] L. P. George, *Automatic mesh generation—Application to Finite Element Methods* (Wiley, New York, 1991).
[14] A. Selamet and P. M. Radavich, ''The effect of length on the acoustic performance of concentric expansion chamber: An analytical, computational and experimental investigation,'' J. Sound Vib. **201**, 407–426 (1997).
[15] H. Kuttruff, *Room Acoustics* (E & FN Spon, New York, 2000).
[16] ASHRAE, *ASHRAE Handbook—Fundamentals* (ASHRAE, Altanta, 2001).
[17] P. O. A. L. Davies, ''Flow-acoustic coupling in ducts,'' J. Sound Vib. **77**, 191–209 (1981).

Acoustic cluster control of noise radiated from a planar structure

Nobuo Tanaka and Ryoji Fukuda
Tokyo Metropolitan Institute of Technology, Department of Mechanical Engineering, 6-6 Asahigaoka, Hino-city, Tokyo 191-0065, Japan

Colin H. Hansen
The University of Adelaide, Department of Mechanical Engineering, Adelaide, SA 5005, Australia

(Received 30 April 2004; revised 19 January 2005; accepted 29 March 2005)

Acoustic cluster control is proposed for the purpose of achieving global sound attenuation of a planar structure. First, acoustic cluster filtering using a point sensor array is presented, which enables the grouping of sound radiated from a target object into a set of clusters, such that each cluster possesses the same common characteristics. This allows the possibility of extracting the cluster of interest without causing observation spillover. Based on the principle of reciprocity, cluster actuation using a point source array is then presented. Driving the source array in accordance with a proposed control law, the excitation of the designated cluster is performed without causing control spillover. Moreover, by combining both acoustic cluster filtering and acoustic cluster actuation, acoustic cluster control may be performed. In implementing acoustic cluster control, the necessary and sufficient condition for the acoustic cluster control is illustrated. It is also shown that the sound radiated from a planar structure may be captured in appropriate acoustic cluster filtering so that acoustic cluster control may be implemented. Experiment was conducted demonstrating the capability as well as the validity of the acoustic cluster filtering, actuation, and control for suppressing the noise radiated from a rectangular panel.
[DOI: 10.1121/1.1920187]

PACS numbers: 43.50.Ki, 43.40.At, 43.40.Dx [KAC] Pages: 3686–3694

I. INTRODUCTION

Free space sound radiation is a problem that has received much attention in the active noise control community. Adaptive feedforward control using a filtered-x least mean square (LMS) algorithm[1] is a commonly reported control strategy. Provided that the system is small and simple, this method may be viable. Once the system becomes large and complicated, however, the control system needs to be large and the computational load on the controller becomes heavy. By increasing the number of sources and sensors, it has been shown that it is possible to increase the area over which attenuation is achieved,[2–5] with significant noise reduction being reported. However, simply attempting to expand source and sensor numbers in an effort to improve global performance is not practically viable; the large number of sensors required as inputs to the control algorithm would quickly overtax the controller central processing unit (CPU) and reduce the performance of the system. The same is true for vibration control strategies. When suppressing the vibration of a distributed parameter structure, control designers face the problem of an immense number of vibration modes, which is theoretically infinite. Of all the vibration control methodologies reported so far, there exists only one control strategy that can deal with an infinite number of structural modes; direct velocity feedback, dubbed DVFB.[6,7] Using collocated velocity sensors and actuators, DVFB enables the damping augmentation of all structural modes; however this method is regarded as low authority control (LAC)[8] as it lacks meticulous control of specific modes of interest.

It is likely that all the structural modes are grouped into a finite number of clusters in which all the structural modes belonging to a particular cluster have the same common characteristics. If the structural modes are orthogonal to each other, the cluster may be controlled independently, thereby enabling the independent control of each cluster using a simple control strategy without causing spillover problems. Grouping all the structural modes into a finite number of clusters is called *cluster filtering*,[9] while independent control of each cluster is termed *cluster actuation*.[9] Utilizing both cluster filtering and cluster actuation, *cluster control*[9] may be performed. This approach falls into a category of middle authority control between conventionally used LAC and high authority control (HAC), possessing the benefits of stability and control law simplicity analogous to LAC, while providing high control performance as well as some flexibility of control gain assignment in a similar way to HAC.

By expanding the conception of cluster control to an active noise control strategy, acoustic cluster control that is able to achieve global suppression of structure-borne noise is proposed. This article begins by presenting acoustic cluster filtering using a point sensor (microphone) array that enables the grouping of the sound radiated from a target object into a set of clusters, such that each cluster possesses the same characteristics in common. As such, it becomes possible to extract the cluster of interest without causing observation spillover. Based on the principle of reciprocity, cluster actuation using a point source (speaker) array is then presented. Driving the source array in accordance with a proposed control law, the excitation of the designated cluster is performed

 0001-4966/2005/117(6)/3686/9/$22.50

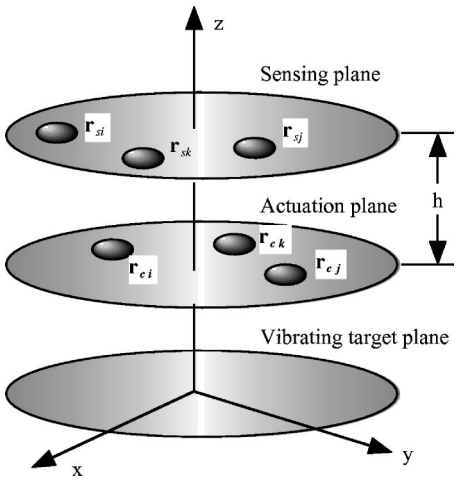


FIG. 1. Geometry of sensing plane, actuation plane, and vibrating target plane.

without causing control spillover. Moreover, by combining both acoustic cluster filtering and acoustic cluster actuation, acoustic cluster control may be performed. The necessary and sufficient condition to implement acoustic cluster control is shown. It is also shown that the sound radiated from a distributed-parameter planar structure may be captured in the appropriate acoustic cluster filtering allowing acoustic cluster control may be implemented. An experiment is carried out to verify the validity of the acoustic cluster filtering and the acoustic cluster actuation. Finally, by employing adaptive feedforward control, acoustic cluster control of the suppression of the noise radiated by a rectangular plate is implemented, demonstrating the capability as well as the validity of the proposed approach.

II. THEORETICAL DEVELOPMENT

A. Acoustic cluster filtering/actuation using point sensors/actuators

As shown in Fig. 1, consider three parallel planes; the top plane is for sensing, the middle for actuation and the bottom for a vibrating target. Suppose that n point sensors (microphones) and n point actuators (speakers) are placed on the sensing plane and actuation plane, respectively, whose locations satisfy the following expressions. The location $\mathbf{r}_{si}$ of the ith point sensor on the sensing plane is written as

$$\mathbf{r}_{si}=(x_{si},y_{si},z_c+h), \tag{1}$$

while the location $\mathbf{r}_{cj}$ of the jth point source (actuator) on the actuation plane is given by

$$\mathbf{r}_{cj}=(x_{cj},y_{cj},z_c). \tag{2}$$

Then, the distance r_{ij} between the ith point sensor (microphone) and the jth point source is

$$r_{ij}=|\mathbf{r}_{si}-\mathbf{r}_{cj}|=\sqrt{(x_{si}-x_{cj})^2+(y_{si}-y_{cj})^2+h^2}. \tag{3}$$

The sound pressure p_i at $\mathbf{r}_{si}$ is expressed as

$$p_i=\frac{j\omega\rho_0}{4\pi}\left(q_1\frac{e^{-jkr_{i1}}}{r_{i1}}+q_2\frac{e^{-jkr_{i2}}}{r_{i2}}+\cdots+q_n\frac{e^{-jkr_{in}}}{r_{in}}\right), \tag{4}$$

where j is the imaginary number, ω is the excitation frequency, ρ_0 is the density of air, q_i is the volume velocity of the ith acoustic source (actuator), and k is the acoustic wave number.

By incorporating all of the acoustic sensor outputs into a vector $\mathbf{p}$, we have

$$\mathbf{p}=\begin{pmatrix}p_1\\p_2\\\vdots\\p_n\end{pmatrix}. \tag{5}$$

Similarly, putting all the volume velocities of the acoustic sources into a vector $\mathbf{q}$, we have

$$\mathbf{q}=\begin{pmatrix}q_1\\q_2\\\vdots\\q_n\end{pmatrix}. \tag{6}$$

Referring to Eq. (4), the relation between the vector $\mathbf{p}$ and $\mathbf{q}$ is given by

$$\mathbf{p}=\mathbf{\Phi q}, \tag{7}$$

where $\mathbf{\Phi}\in\mathbf{C}^{n\times n}$ denotes the acoustic impedance matrix defined as

$$\mathbf{\Phi}=\frac{j\omega\rho_0}{4\pi}\begin{pmatrix}\frac{e^{-jkr_{11}}}{r_{11}} & \frac{e^{-jkr_{12}}}{r_{12}} & \cdots & \frac{e^{-jkr_{1n}}}{r_{1n}}\\ \frac{e^{-jkr_{21}}}{r_{21}} & \frac{e^{-jkr_{22}}}{r_{22}} & \cdots & \frac{e^{-jkr_{2n}}}{r_{2n}}\\ \cdot & \cdot & \cdot & \cdot\\ \frac{e^{-jkr_{n1}}}{r_{n1}} & \frac{e^{-jkr_{n2}}}{r_{n2}} & \cdots & \frac{e^{-jkr_{nn}}}{r_{nn}}\end{pmatrix}. \tag{8}$$

Next, using a nonsingular matrix $\mathbf{T}\in\mathbf{R}^{n\times n}$, a cluster filtering is performed as

$$\tilde{\mathbf{p}}=\mathbf{Tp}. \tag{9}$$

Note that the ith element $\tilde{p}_i$ of the vector $\tilde{\mathbf{p}}$ is expressed as the combination of all the elements of the vector $\mathbf{p}$ with the weighting factors, $t_{ij}\,(j=1\sim n)$, the ith and jth element of $\mathbf{T}$, thus allowing the extraction of some information out of the sensor outputs. In exactly the same way, the sound cluster actuation may be carried out by

$$\tilde{\mathbf{q}}=\mathbf{Tq}. \tag{10}$$

Note again that the ith element $\tilde{q}_i$ of the vector $\tilde{\mathbf{q}}$ is expressed as the combination of all the elements of the vector $\mathbf{q}$ with the same weighting factors, $t_{ij}\,(j=1\sim n)$, used for the sound cluster filtering. Using Eqs. (7), (9), and (10), we have

$$\tilde{\mathbf{p}}=\mathbf{Tp}=\mathbf{T\Phi T}^{-1}\tilde{\mathbf{q}}. \tag{11}$$

Suppose that Eq. (11) is expressed as

$$\tilde{\mathbf{p}}=\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & 0 & 0\\ \vdots & 0 & \ddots & \\ 0 & 0 & 0 & \lambda_n\end{pmatrix}\tilde{\mathbf{q}}. \tag{12}$$

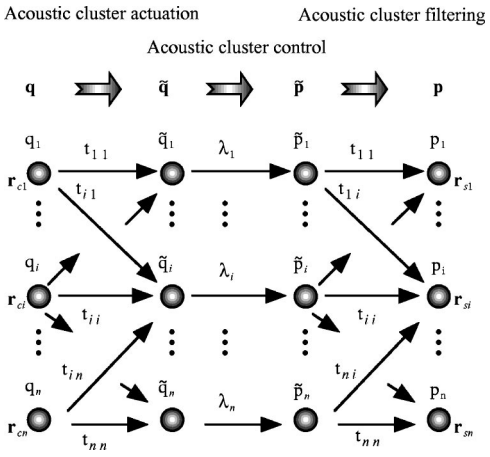


FIG. 2. Signal flow of acoustic cluster control with acoustic cluster filtering and actuation.

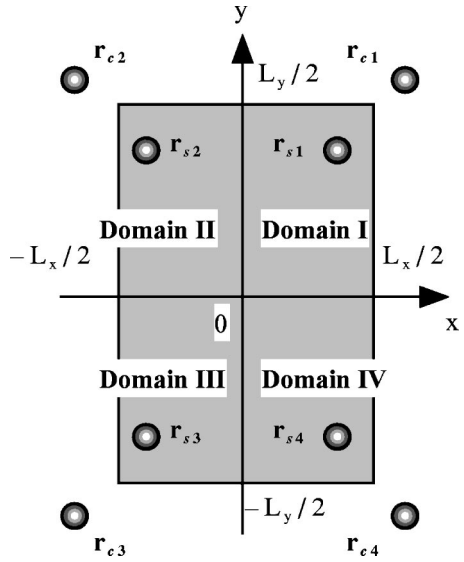


FIG. 3. Coordinate system of a rectangular plate showing the configuration of sources and sensors.

Then, sound cluster control, achieved as the ith element $\tilde{p}_i$ of $\tilde{\mathbf{p}}$, for instance, is influenced by only the ith element $\tilde{q}_i$ of the vector $\tilde{\mathbf{q}}$, and so independent control may be performed without causing control/observation spillover. Equation (12) is thus termed *a cluster equation*. As such, in order to implement acoustic cluster control, the acoustic impedance matrix $\mathbf{\Phi}$ needs to be diagonalized. If $\mathbf{\Phi}$ were a real symmetrical matrix, it could be diagonalized using an orthonormal matrix. However $\mathbf{\Phi}$ is a complex matrix, so that its diagonalization is not straightforward. To do so, first, it needs to be a "normal matrix."[10] If $\mathbf{\Phi}$ is a normal matrix, then it can be diagonalized by using an appropriate unitary matrix. The necessary and sufficient condition for $\mathbf{\Phi}$ to be a normal matrix is to satisfy the following condition;

$$\mathbf{\Phi}^H\mathbf{\Phi}=\mathbf{\Phi}\mathbf{\Phi}^H, \tag{13}$$

where H denotes the matrix Hermitian (conjugate transpose).

A signal flow diagram of the cluster control is depicted in Fig. 2. Observe that there is no cross talk between the acoustic cluster filtering set $\tilde{p}_i$ and the acoustic cluster actuation set $\tilde{q}_i$; $\tilde{q}_i$ is directly transmitted to $\tilde{p}_i$ with the gain λ_i without causing any spillover. Here, λ_i is the eigenvalue of the matrix $\mathbf{\Phi}$, playing the role of a control gain on $\tilde{q}_i$, and so it may be adjusted in an effort to meet the desired weight on each modal cluster.

Using an example, it is worthwhile to show how the acoustic cluster filtering and actuation work. Assume the acoustic point sources at $\mathbf{r}_{c1}\sim\mathbf{r}_{c4}$ and acoustic point sensors at $\mathbf{r}_{s1}\sim\mathbf{r}_{s4}$ are placed, as shown in Fig. 3, on the sensing plane and actuation plane, respectively, such that the following descriptions are satisfied,

$$\mathbf{r}_{c1}=(x_c,y_c,z_c), \tag{14}$$

$$\mathbf{r}_{c2}=(-x_c,y_c,z_c), \tag{15}$$

$$\mathbf{r}_{c3}=(-x_c,-y_c,z_c), \tag{16}$$

$$\mathbf{r}_{c4}=(x_c,-y_c,z_c), \tag{17}$$

$$\mathbf{r}_{s1}=(x_s,y_s,z_c+h), \tag{18}$$

$$\mathbf{r}_{s2}=(-x_s,y_s,z_c+h), \tag{19}$$

$$\mathbf{r}_{s3}=(-x_s,-y_s,z_c+h), \tag{20}$$

$$\mathbf{r}_{s4}=(x_s,-y_s,z_c+h). \tag{21}$$

Referring to the configuration of four point sensors and four point actuators as shown in Fig. 3, the impedance matrix in Eq. (8) is given by

$$\mathbf{\Phi}=\begin{pmatrix} a & b & c & d \\ b & a & d & c \\ c & d & a & b \\ d & c & b & a \end{pmatrix}, \tag{22}$$

where

$$a=\frac{j\omega\rho_0 e^{-jkr_{11}}}{4\pi r_{11}}, \tag{23}$$

$$b=\frac{j\omega\rho_0 e^{-jkr_{12}}}{4\pi r_{12}}, \tag{24}$$

$$c=\frac{j\omega\rho_0 e^{-jkr_{13}}}{4\pi r_{13}}, \tag{25}$$

$$d=\frac{j\omega\rho_0 e^{-jkr_{14}}}{4\pi r_{14}}. \tag{26}$$

Note that the acoustic impedance matrix in Eq. (22) satisfies the necessary and sufficient condition in Eq. (13), so that it is a normal matrix, and hence diagonalized as described shortly. The eigenvalues of $\mathbf{\Phi}$ are calculated as

$$\lambda_1=a+b+c+d, \tag{27}$$

$$\lambda_2=a+b-c-d, \tag{28}$$

$$\lambda_3=a-b-c+d, \tag{29}$$

$$\lambda_4=a-b+c-d. \tag{30}$$

The modal matrix $\mathbf{T}$ associated with the eigenvalues in Eqs. (27)–(30) yields

$$\mathbf{T}=\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & -1 & 1\\ 1 & -1 & 1 & -1\end{pmatrix}. \tag{31}$$

Note again that the matrix $\mathbf{T}$ is a Hadamard matrix as was previously used in the study of vibration transmission,[11,12] and is a special case of a unitary matrix. The cluster equation in Eq. (12) is then written as

$$\tilde{\mathbf{p}}=\begin{pmatrix}\lambda_1 & 0 & 0 & 0\\ 0 & \lambda_2 & 0 & 0\\ 0 & 0 & \lambda_3 & 0\\ 0 & 0 & 0 & \lambda_4\end{pmatrix}\tilde{\mathbf{q}}. \tag{32}$$

Clearly, from Eq. (32), cluster control can be achieved with the arrangement of actuators and sensors as shown in Fig. 3. For example, consider the first monopole pattern $\tilde{q}_1$ that is obtained by adding up all the acoustic source outputs in phase. This information may be extracted in the cluster filtering $\tilde{p}_1$ that is implemented by summing up all the sensor outputs in phase, yielding

$$\tilde{p}_1=\lambda_1\tilde{q}_1=\lambda_1(q_1+q_2+q_3+q_4). \tag{33}$$

However, the information—the transformed source set $\tilde{q}_1$—does not emerge in the other cluster filtering outputs, and hence the outputs of $\tilde{p}_2$, $\tilde{p}_3$, and $\tilde{p}_4$ in terms of $\tilde{q}_1$ are zero.

B. Acoustic cluster filtering/actuation of a rectangular plate

1. Acoustic field function

The sound pressure at $\mathbf{r}_k$ in space induced by a baffled rectangular plate with the dimension of $L_x\times L_y\times t$, subject to a harmonic excitation, is given by

$$p(\mathbf{r}_k)=\int_D v(\mathbf{r}_p)z_p(\mathbf{r}_p|\mathbf{r}_k)d\mathbf{r}_p, \tag{34}$$

where $z_p(\mathbf{r}_p|\mathbf{r}_k)$ is the acoustic impedance from $\mathbf{r}_p$ on the target plate to $\mathbf{r}_k$ in space, $v(\mathbf{r}_p)$ is the velocity at $\mathbf{r}_p$, and D is the domain of the plate. Furthermore, the velocity $v(\mathbf{r}_p)$ may be described using the expansion theory as

$$v(\mathbf{r}_p)=\sum_{i=1}^{\infty}\varphi_i(\mathbf{r}_p)v_i, \tag{35}$$

where $\varphi_i(\mathbf{r}_p)$ is the ith eigenfunction of the panel and v_i is the associated modal amplitude. By inserting Eq. (35) into Eq. (34), the sound pressure is given by

$$\begin{aligned}p(\mathbf{r}_k)&=\int_D\sum_{i=1}^{\infty}\varphi_i(\mathbf{r}_p)v_iz_p(\mathbf{r}_p|\mathbf{r}_k)d\mathbf{r}_p\\ &=\sum_{i=1}^{\infty}v_i\int_D\varphi_i(\mathbf{r}_p)z_p(\mathbf{r}_p|\mathbf{r}_k)d\mathbf{r}_p.\end{aligned} \tag{36}$$

As shown in Fig. 3, the target plate is divided into four domains: I through IV. In each domain of the plate, a pair of point acoustic sensors and actuators are placed as expressed in Eqs. (14)–(21). By combining the four sound pressures into the vector $\mathbf{p}$, we have

$$\mathbf{p}=\begin{pmatrix}p(\mathbf{r}_{s1})\\ p(\mathbf{r}_{s2})\\ p(\mathbf{r}_{s3})\\ p(\mathbf{r}_{s4})\end{pmatrix}=\begin{pmatrix}p_1\\ p_2\\ p_3\\ p_4\end{pmatrix}. \tag{37}$$

Moreover, by introducing the Hadamard matrix in Eq. (31), the sound pressure vector $\mathbf{p}$ is transformed into

$$\tilde{\mathbf{p}}=\begin{pmatrix}\tilde{p}_1\\ \tilde{p}_2\\ \tilde{p}_3\\ \tilde{p}_4\end{pmatrix}=\mathbf{T}\mathbf{p}. \tag{38}$$

Thus, the elements of $\tilde{\mathbf{p}}$ in Eq. (38) are expressed by addition/subtraction processing of sound pressure outputs owing to the property of the Hadamard matrix in Eq. (31), and further expand to

$$\tilde{p}_1=\sum_{i=1}^{\infty}v_i\int_D\varphi_i(\mathbf{r}_p)\psi^1(\mathbf{r}_p)d\mathbf{r}_p, \tag{39}$$

$$\tilde{p}_2=\sum_{i=1}^{\infty}v_i\int_D\varphi_i(\mathbf{r}_p)\psi^2(\mathbf{r}_p)d\mathbf{r}_p, \tag{40}$$

$$\tilde{p}_3=\sum_{i=1}^{\infty}v_i\int_D\varphi_i(\mathbf{r}_p)\psi^3(\mathbf{r}_p)d\mathbf{r}_p, \tag{41}$$

$$\tilde{p}_4=\sum_{i=1}^{\infty}v_i\int_D\varphi_i(\mathbf{r}_p)\psi^4(\mathbf{r}_p)d\mathbf{r}_p, \tag{42}$$

where $\psi^i(\mathbf{r}_p)$ denotes the acoustic filed function defined as

$$\begin{aligned}\psi^1(\mathbf{r}_p)&=z_p(\mathbf{r}_p|\mathbf{r}_{s1})+z_p(\mathbf{r}_p|\mathbf{r}_{s2})+z_p(\mathbf{r}_p|\mathbf{r}_{s3})\\ &\quad+z_p(\mathbf{r}_p|\mathbf{r}_{s4}),\end{aligned} \tag{43}$$

$$\begin{aligned}\psi^2(\mathbf{r}_p)&=z_p(\mathbf{r}_p|\mathbf{r}_{s1})+z_p(\mathbf{r}_p|\mathbf{r}_{s2})-z_p(\mathbf{r}_p|\mathbf{r}_{s3})\\ &\quad-z_p(\mathbf{r}_p|\mathbf{r}_{s4}),\end{aligned} \tag{44}$$

$$\begin{aligned}\psi^3(\mathbf{r}_p)&=z_p(\mathbf{r}_p|\mathbf{r}_{s1})-z_p(\mathbf{r}_p|\mathbf{r}_{s2})-z_p(\mathbf{r}_p|\mathbf{r}_{s3})\\ &\quad+z_p(\mathbf{r}_p|\mathbf{r}_{s4}),\end{aligned} \tag{45}$$

$$\begin{aligned}\psi^4(\mathbf{r}_p)&=z_p(\mathbf{r}_p|\mathbf{r}_{s1})-z_p(\mathbf{r}_p|\mathbf{r}_{s2})+z_p(\mathbf{r}_p|\mathbf{r}_{s3})\\ &\quad-z_p(\mathbf{r}_p|\mathbf{r}_{s4}),\end{aligned} \tag{46}$$

and where the acoustic impedance $z_p(\mathbf{r}_p|\mathbf{r}_{si})$ is written as

$$z_p(\mathbf{r}_p|\mathbf{r}_{si})=\frac{j\omega\rho_0}{2\pi r_i}e^{-jkr_i}. \tag{47}$$

Then, the distance $r_i\,(i=1-4)$ between $\mathbf{r}_p=(x_p,y_p)$, the arbitrary position of the panel, and $\mathbf{r}_{si}$, the sensor location as given by Eqs. (18)–(21), yields

$$r_1=\sqrt{(x_p-x_s)^2+(y_p-y_s)^2+(z_c+h)^2}, \tag{48}$$

$$r_2=\sqrt{(x_p+x_s)^2+(y_p-y_s)^2+(z_c+h)^2}, \tag{49}$$

$$r_3=\sqrt{(x_p+x_s)^2+(y_p+y_s)^2+(z_c+h)^2}, \tag{50}$$

$$r_4=\sqrt{(x_p-x_s)^2+(y_p+y_s)^2+(z_c+h)^2}. \tag{51}$$

2. Fundamental characteristics of the field function

Consider the field functions expressed in Eqs. (43)–(46). Of four acoustic impedance terms; $z_p(\mathbf{r}_p|\mathbf{r}_{s1})$, $z_p(\mathbf{r}_p|\mathbf{r}_{s2})$, $z_p(\mathbf{r}_p|\mathbf{r}_{s3})$, and $z_p(\mathbf{r}_p|\mathbf{r}_{s4})$ comprising the field functions, let us pick up, say $z_p(\mathbf{r}_p|\mathbf{r}_{s1})$, for investigating the fundamental characteristics of the field functions. The real and imaginary parts of $z_p(\mathbf{r}_p|\mathbf{r}_{s1})$, respectively, yield

$$\mathrm{Re}[z_p(\mathbf{r}_p|\mathbf{r}_{s1})]=\frac{\omega\rho_0 k}{2\pi}\mathrm{sinc}\,k\mathbf{r}_1, \tag{52}$$

and

$$\mathrm{Im}[z_p(\mathbf{r}_p|\mathbf{r}_{s1})]=\frac{\omega\rho_0}{2\pi r_1}\cos kr_1. \tag{53}$$

Clearly, from Eqs. (52) and (53), both of the functions have a maximum, $\omega\rho_0 k/2\pi_1\mathrm{sinc}\,k(z_c+h)$ and $\omega\rho_0/2\pi k(z_c+h)\cos k(z_c+h)$, respectively, at $\mathbf{r}_p=(x_s,y_s)$, with the values of these terms decreasing monotonously along the x and y axes. Other impedance functions also entail the same characteristics, possessing maxima at $\mathbf{r}_p=(-x_s,y_s)$, $\mathbf{r}_p=(-x_s,-y_s)$, and $\mathbf{r}_p=(x_s,-y_s)$, respectively.

The fundamental characteristics of the field function characterized by the impedance terms are thus summarized as:

(1) Field function; $\psi^1(\mathbf{r})$,
Even function with respect to both the x and y axes, and hence

$$\psi^1(x_1,y_1)=\psi^1(-x_1,y_1)=\psi^1(-x_1,-y_1)=\psi^1(x_1,-y_1). \tag{54}$$

(2) Field function; $\psi^2(\mathbf{r})$,
Even and odd function with respect to the x and y axes, respectively, and hence

$$\psi^2(x_1,y_1)=\psi^2(-x_1,y_1)=-\psi^2(-x_1,-y_1)=-\psi^2(x_1,-y_1). \tag{55}$$

(3) Field function; $\psi^3(\mathbf{r})$,
Odd and even function with respect to the x and y axes, respectively, and hence

$$\psi^3(x_1,y_1)=-\psi^3(-x_1,y_1)=-\psi^3(-x_1,-y_1)=\psi^3(x_1,-y_1). \tag{56}$$

(4) Field function; $\psi^4(\mathbf{r})$,
Odd function with respect to both the x and y axes, and hence

$$\psi^4(x_1,y_1)=-\psi^4(-x_1,y_1)=\psi^4(-x_1,-y_1)=-\psi^4(x_1,-y_1). \tag{57}$$

3. Acoustic cluster filtering mechanism

Consider the case where all the modal functions, $\varphi(\mathbf{r})$, of a rectangular plate are classified into four categories with the characteristics described as follows:

(1) Odd/odd modal function; $\varphi^{o/o}(\mathbf{r})$,

$$\varphi^{o/o}(x_1,y_1)=\varphi^{o/o}(-x_1,y_1)=\varphi^{o/o}(-x_1,-y_1)=\varphi^{o/o}(x_1,-y_1), \tag{58}$$

where the superscript denotes that the numbers of nodal areas of a panel in the x and y directions are odd and odd, respectively. Note that the odd/odd modal functions comprise the first acoustic power mode[13,14] which is the greatest contributor to the total acoustic power radiated from a rectangular plate.

(2) Odd/even modal function; $\varphi^{o/e}(\mathbf{r})$,

$$\varphi^{o/e}(x_1,y_1)=\varphi^{o/e}(-x_1,y_1)=-\varphi^{o/e}(-x_1,-y_1)=-\varphi^{o/e}(x_1,-y_1). \tag{59}$$

(3) Even/odd modal function; $\varphi^{e/o}(\mathbf{r})$,

$$\varphi^{e/o}(x_1,y_1)=-\varphi^{e/o}(-x_1,y_1)=-\varphi^{e/o}(-x_1,-y_1)=\varphi^{e/o}(x_1,-y_1). \tag{60}$$

(4) Even/even modal function; $\varphi^{e/e}(\mathbf{r})$,

$$\varphi^{e/e}(x_1,y_1)=-\varphi^{e/e}(-x_1,y_1)=\varphi^{e/e}(-x_1,-y_1)=-\varphi^{e/e}(x_1,-y_1). \tag{61}$$

Boundary conditions satisfying the above properties are, for instance, simple support of four edges of the plate (SSSS), clamped support (CCCC), free support (FFFF), and combinations thereof such as (SCSC), (FCFC), and (SFSF), where an eigenfunction is expressed as a symmetrical or skew-symmetrical function with respect to the x and y axes.

If all the modal functions of the rectangular panel are classified into the aforementioned categories, then acoustic cluster filtering may be achieved as expressed below. To do so, consider, for instance, the acoustic cluster filtering output $\tilde{p}_1$ in Eq. (39). In order to perform the required integrations, the target domain is divided into four subdomains as shown in Fig. 3, and so the integration is also divided into four terms:

$$\begin{aligned}\tilde{p}_1&=\sum_{i=1}^{\infty}v_i\int_D\varphi_i(\mathbf{r}_p)\psi^1(\mathbf{r}_p)d\mathbf{r}_p,\\&=\sum_{i=1}^{\infty}v_i\int_0^{Lx/2}\int_0^{Lx/2}\varphi_i(x,y)\psi^1(x,y)dxdy,\\&\quad+\sum_{i=1}^{\infty}v_i\int_{-Lx/2}^{0}\int_0^{Lx/2}\varphi_i(x,y)\psi^1(x,y)dxdy,\\&\quad+\sum_{i=1}^{\infty}v_i\int_{-Lx/2}^{0}\int_{-Lx/2}^{0}\varphi_i(x,y)\psi^1(x,y)dxdy,\\&\quad+\sum_{i=1}^{\infty}v_i\int_0^{Lx/2}\int_{-Lx/2}^{0}\varphi_i(x,y)\psi^1(x,y)dxdy.\end{aligned} \tag{62}$$

Of the four clusters of modal functions, consider, for instance, the odd/odd modal function, $\varphi^{o/o}(x,y)$. Then, due to the common characteristics of both the modal function $\varphi^{o/o}(x,y)$ and $\psi^1(x,y)$, which are even functions with respect to the x and y axes, Eq. (62) further expands to

$$\tilde{p}_1 = \sum_{i=1}^{\infty} v_i \int_D \varphi_i^{o/o}(\mathbf{r}_p)\psi^1(\mathbf{r}_p)d\mathbf{r}_p,$$

$$= 4\sum_{i=1}^{\infty} v_i \int_0^{Lx/2}\int_0^{Lx/2} \varphi_i^{o/o}\psi^1(x,y)dxdy, \quad (63)$$

where $\tilde{p}_1$ is nonzero. On the contrary, Eq. (62) in terms of the other three modal functions becomes

$$\tilde{p}_1 = \sum_{i=1}^{\infty} v_i \int_D \varphi_i^{o/e}(\mathbf{r}_p)\psi^1(\mathbf{r}_p)d\mathbf{r}_p = 0, \quad (64)$$

$$\tilde{p}_1 = \sum_{i=1}^{\infty} v_i \int_D \varphi_i^{e/o}(\mathbf{r}_p)\psi^1(\mathbf{r}_p)d\mathbf{r}_p = 0, \quad (65)$$

$$\tilde{p}_1 = \sum_{i=1}^{\infty} v_i \int_D \varphi_i^{e/e}(\mathbf{r}_p)\psi^1(\mathbf{r}_p)d\mathbf{r}_p = 0. \quad (66)$$

Thus, the modal function, the odd/odd modal function in this case, is orthogonal with respect to the field function, $\psi^1(x,y)$, thereby enabling acoustic cluster filtering in the sense of cluster dimension. As a result, the acoustic cluster output $\tilde{p}_1$ consists of all the odd/odd modal functions, with no other modal functions involved. Similarly, the fundamental characteristics of the other cluster filterings are investigated and concluded as follows: $\tilde{p}_2$, $\tilde{p}_3$, and $\tilde{p}_4$ permit the acoustic cluster filtering of the odd/even modal functions, even/odd modal clusters and even/even modal clusters, respectively.

III. NUMERICAL SIMULATION

To clarify the validity of the proposed acoustic cluster filtering, cluster actuation, and cluster control, numerical simulations were conducted. A simply supported, rectangular, baffled, steel panel with the dimensions $1.2\,\mathrm{m}\,(L_x) \times 0.6\,\mathrm{m}\,(L_y) \times 0.004\,\mathrm{m}\,(t)$ is used as a target plate excited at $\mathbf{r}_d = (0.256\,\mathrm{m}, 0.356\,\mathrm{m})$ by a disturbance point force. Four microphones for cluster filtering are placed at $x_s = y_s = z_s = 10\,\mathrm{cm}$, hence $\mathbf{r}_{s1}$, $\mathbf{r}_{s2}$, $\mathbf{r}_{s3}$, and $\mathbf{r}_{s4}$ may be determined according to Eqs. (18)–(21). Figure 4(a) shows the frequency characteristics of the sound pressure at $\mathbf{r}_{s1}$, and seven resonant peaks appear in the frequency range up to 200 Hz. Acoustic cluster filtering effects are shown in Figs. 4(b)–4(e); referring to Fig. 4(b), no resonant peaks except the target modes, the (1,1) and (1,3) modes, appear in the acoustic cluster $\tilde{p}_1$. Figure 4(c) illustrates the frequency characteristics of the acoustic cluster filtering output $\tilde{p}_2$. Note that, referring to Table I listing the modal number and frequencies of the plate, there is a pair of ''almost degenerate structural modes''; {(1,4) mode and (2,2) mode}, the resonance frequencies of which are very close to each other. Clearly from Figs. 4(c) and 4(e), the (1,4) mode and the (2,2) mode are decoupled and extracted in the acoustic clusters $\tilde{p}_2$ and $\tilde{p}_4$, respectively. As with the acoustic cluster $\tilde{p}_3$, resonant peaks with respect to the (2,1) and the (2,3) mode appear in Fig. 4(d). As such, out of the structural modes of interest only the designated modes are perfectly extracted in the acoustic cluster filtering.

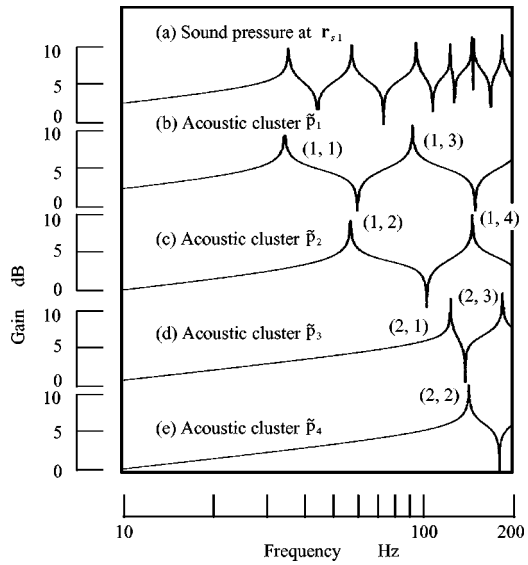


FIG. 4. Frequency characteristics of sound pressure at $\mathbf{r}_{s1}$ and acoustic cluster outputs.

It is worthwhile to clarify the limitations of acoustic cluster control with reference to three particular questions as discussed in the following paragraphs.

(1) *What is the maximum frequency that cluster control may deal with*? This question is to be raised regardless of cluster control or conventional active noise control. In practice, it depends on the performance of the actuators and digital signal processes (DSPs) used for the controller. Theoretically, the wavelength of the maximum frequency of interest should not be much smaller than a structural dimension. As the panel becomes larger, the number of actuators needs to be increased to achieve the desired control effect.

(2) *Computation load*; Simply attempting to increase the actuator and sensor numbers in an effort to improve global performance of a traditional active noise control system is not practically viable; as the number of sensors required increases, the controller performance reduces and the CPU soon becomes incapable of handling the load. To cope with this problem, cluster control is suggested; a multi-input multi-output (MIMO) system used for active noise control (ANC) may be replaced with a finite number of single input single outputs (SISOs), thus significantly reducing the required computation load. Even in

TABLE I. Mode number, modal indices, and modal resonance frequencies of the target rectangular panel.

Mode number	Mode	Modal Frequency Hz
1	(1,1)	34.49
2	(1,2)	55.51
3	(1,3)	90.55
4	(2,1)	116.94
5	(2,2)	137.96
6	(1,4)	139.60
7	(2,3)	173.00

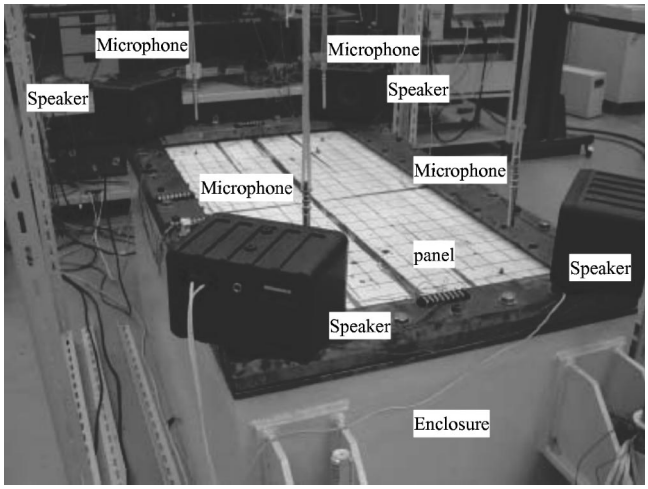


FIG. 5. Experimental arrangement showing the panel, speakers, and microphones.

the worst case, when the cluster control fails to be implemented, the cluster control system becomes a conventional MIMO system.

(3) *Arbitrary shape of a planar structure*; Since a simply supported rectangular panel was used as an example, a question may arise; how is it possible to extend the work to an arbitrarily shaped planar structure. This problem may be tackled by replacing the vibrating structure, irrespective of its size, shape or complexity, with a virtual sound source array on a virtual plane[15] in the vicinity of the target structure. If this is the case, virtual sound sources may be allocated methodically on a simplified virtual plane such as a simple rectangular panel as was used in the current work. With the virtual sound source array, cluster control consisting of cluster filtering and cluster actuation can be applied in an effort to achieve a global suppression of the structure-borne sound radiated from a planar structure.

IV. EXPERIMENT

With a view to verifying the validity of the proposed method, an experiment was carried out. As shown in Fig. 5, a steel rectangular panel with the same dimensions as used in the numerical simulation was fixed on top of an enclosure constructed from 10 cm thick concrete walls. The uppermost edges of the walls consist of a ''knife blade'' in order to simulate simply supported boundary conditions. The enclosed cavity was packed with glasswool to dampen acoustic resonances. An electrodynamic shaker attached at $\mathbf{r}_d = (0.256\,\mathrm{m}, 0.356\,\mathrm{m})$, the same location used in simulation, provided the panel with a random disturbance force, low-pass filtered at 250 Hz. Four microphones used for acoustic cluster filtering were hung from a scaffolding placed around the test rig. One microphone was situated at $\mathbf{r}_{s1} = (0.1\,\mathrm{m}, 0.4\,\mathrm{m}, 0.1\,\mathrm{m})$, and so the other locations $\mathbf{r}_{s2} \sim \mathbf{r}_{s4}$ were determined as shown in Eqs. (18)–(21).

A. Acoustic cluster filtering

Figures 6(a) and 6(b) show the driving point mobility of the panel at $\mathbf{r}_d$, and the associated frequency spectrum of the sound pressure at $\mathbf{r}_{s1}$, respectively. Figures 6(c)–6(f) depict the frequency characteristics of the acoustic cluster filtering outputs, $\tilde{p}_1$, $\tilde{p}_2$, $\tilde{p}_3$, and $\tilde{p}_4$, respectively. As seen from the figures, seven structural modes under consideration were clustered into the appropriate cluster filtering output. As seen in the simulation results, the (1,4) mode and the (2,2) mode; ''almost degenerate modes'', were extracted experimentally in the acoustic cluster $\tilde{p}_2$ and $\tilde{p}_4$, respectively. Please note that the experiment was carried out in an ordinary room, and not under ideal circumstances such as found in an anechoic chamber. However, even under such a poor condition, the experimental data show satisfactory results. Although the use of microphones has the shortcoming of being vulnerable to environmental noise, the disturbance noise may be cancelled in the acoustic cluster filtering, $\tilde{p}_2$, $\tilde{p}_3$, and $\tilde{p}_4$ because of the subtraction operation in the signal processing, given that the same noise enters the four microphones simultaneously.

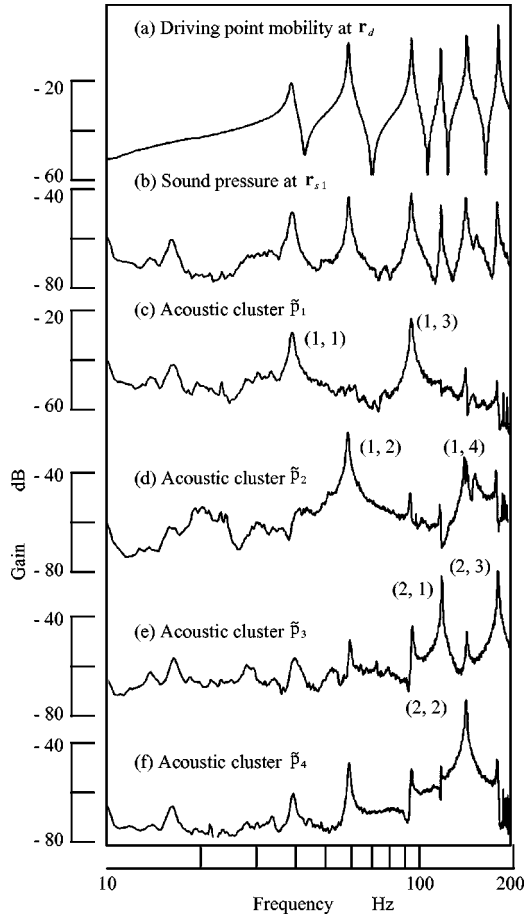


FIG. 6. Frequency characteristics of: (a) driving point mobility at $\mathbf{r}_d$ of the panel, (b) sound pressure at $\mathbf{r}_{s1}$, and (c)–(f) acoustic cluster outputs.

B. Acoustic cluster actuation

One speaker was placed at $\mathbf{r}_{c1} = (0.3\,\mathrm{m}, 0.55\,\mathrm{m}, 0.05\,\mathrm{m})$ and the three other locations were determined according to Eqs. (14)–(17). To verify the validity of the acoustic cluster actuation, an experiment was conducted. By driving four speakers at 100 Hz in phase, the cluster actuation signal $\tilde{q}_1$ was produced [see Eq. (10)]. If this clustered sound $\tilde{q}_1$ is captured *only* in the acoustic cluster filtering output $\tilde{p}_1$ and does not appear in the other acoustic cluster filtering outputs,

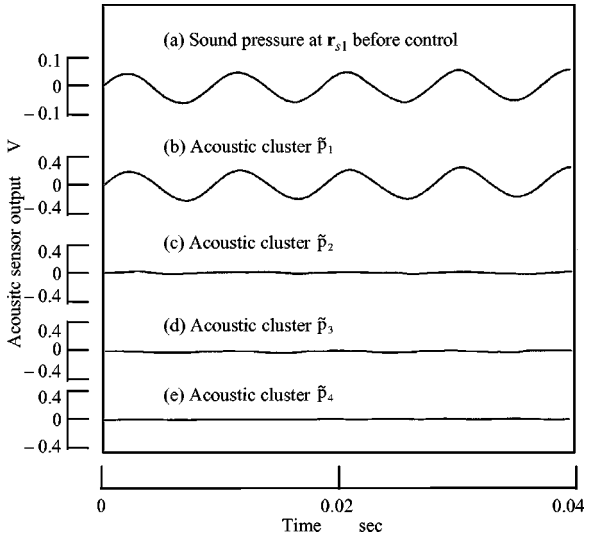


FIG. 7. Time histories of: (a) sound pressure at $\mathbf{r}_{s1}$, and (b)–(e) acoustic cluster outputs.

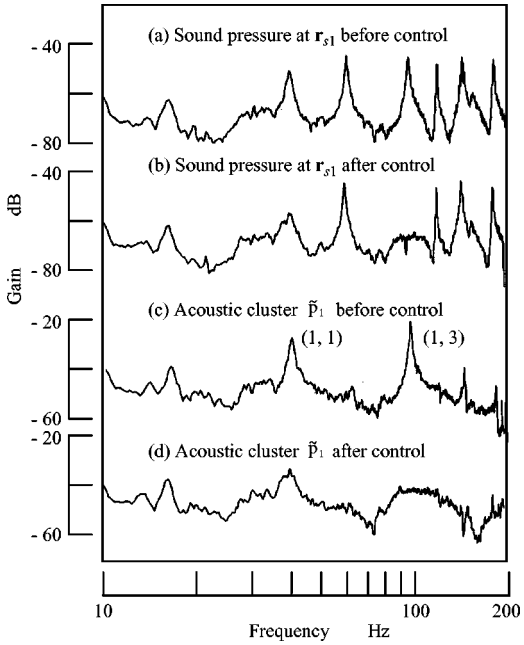


FIG. 9. Frequency characteristics of: (a) sound pressure at $\mathbf{r}_{s1}$ before control, (b) sound pressure at $\mathbf{r}_{s1}$ after control, (c) acoustic cluster $\tilde{p}_1$ before control, and (d) acoustic cluster $\tilde{p}_1$ after control.

then this represents proof of the acoustic cluster actuation. Figure 7(a) shows the time history of the sound pressure measured at $\mathbf{r}_{s1}$, while Figs. 7(b)–7(e) illustrate those of the acoustic cluster, $\tilde{p}_1$, $\tilde{p}_2$, $\tilde{p}_3$, and $\tilde{p}_4$, respectively. As seen from the figures, only the acoustic cluster $\tilde{p}_1$ extracted the cluster sound $\tilde{q}_1$, whereas the other acoustic clusters did not. The amplitude of the sound pressure in Fig. 7(b) is four times as large as that in Fig. 7(a) since the summation operation in the signal processing is involved.

C. Acoustic cluster control

Figure 8(a) illustrates the schematic diagram of a conventional active noise control system with four inputs and four outputs. With a view to achieving global noise suppression, it is required to increase the number of sensors and speakers. However, the large number of sensors required as input to the control algorithm would quickly overtax the controller CPU and reduce the performance of the system even with the ever-increasing capabilities of microprocessors. On the contrary, the acoustic cluster control presented in the current work enables the independent control of each cluster, thereby reducing the computational burden. More importantly, the cluster filtering output is no longer the sound pressure at a sensor location, but the set of all structural modes whose radiation is responsible for the total acoustic power radiated by the target structure. Thus, the suppression of the cluster filtering outputs leads to the suppression of the global noise attenuation. Furthermore, a four input four output control system is regarded as four SISO systems in Fig. 8(b) as opposed to MIMO in Fig. 8(a), thus reducing the computation load.

Using feedforward control based upon a filtered-x LMS algorithm, acoustic cluster control consisting of both acoustic cluster filtering using four microphones and acoustic cluster actuation using four speakers, was implemented in the experiment. Four speakers were driven in accordance with the cluster actuation law stated previously. Control signals driving four speakers were obtained by addition/subtraction of the control signal outputs; however they are *not* cancelled as the control signals consist of eigenfunctions of the target

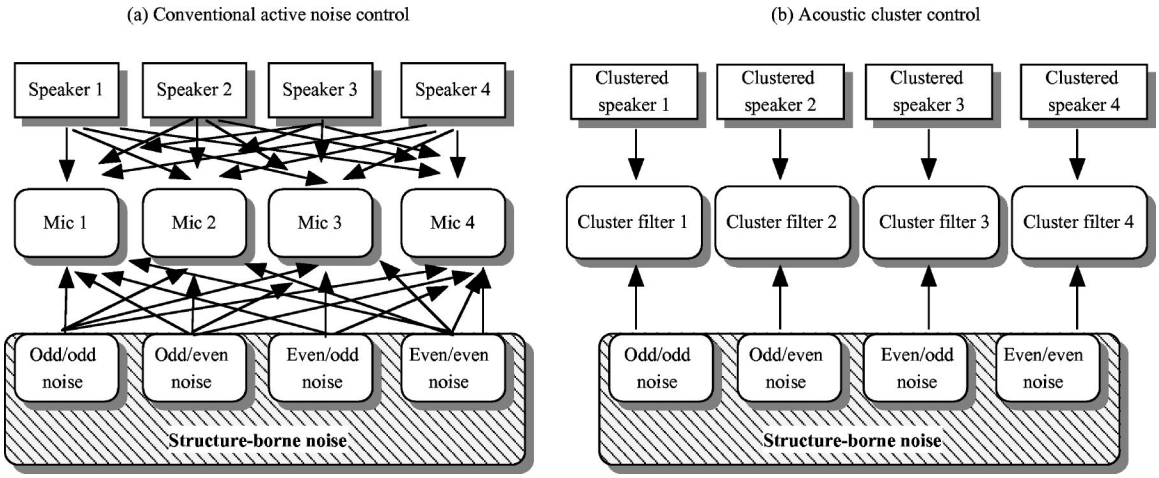


FIG. 8. Schematic diagram of (a) conventional active noise control and (b) acoustic cluster control.

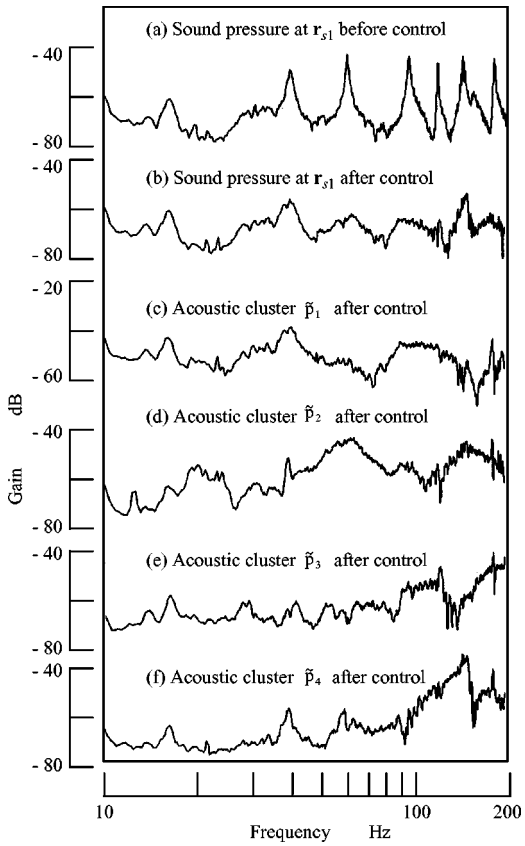


FIG. 10. Frequency characteristics of: (a) sound pressure at $\mathbf{r}_{s1}$ before control, (b) sound pressure at $\mathbf{r}_{s1}$ after control, and (c)–(f) acoustic cluster output after control.

structure, which are orthogonal to each other.

Figure 9 shows the case when only the acoustic cluster $\tilde{p}_1$ was controlled, whereas the other three were not. Figures 9(a) and 9(b) illustrate the frequency characteristics of the sound pressure at $\mathbf{r}_{s1}$ before and after cluster control. In the frequency range under consideration, seven resonant peaks should exist in Fig. 9(a), although only six resonant peaks are seen because of a pair of almost degenerate modes. The resonant peaks of the target modes, the (1,1) and (1,3) modes, are suppressed as seen in Fig. 9(b), while the other peaks remained unaffected. It may be seen that the resonant peaks of the (1,1) and (1,3) mode were reduced by 9 dB and 23 dB, respectively, whereas the other modal resonant peaks remained intact. Figures 9(c) and 9(d) show the frequency characteristics of the acoustic cluster output $\tilde{p}_1$ before and after cluster control, respectively. As seen in Figs. 9(b) and 9(d), only the target modal clusters were suppressed.

In an effort to suppress all four acoustic cluster outputs, a further experiment was undertaken. Suppression of the four clusters is equivalent to the suppression of all the resonant peaks of the target plate. Figures 10(a) and 10(b) show the frequency characteristics of the sound pressure measured at $\mathbf{r}_{s1}$ before and after acoustic cluster control. As seen from Fig. 10(b), all seven resonant peaks were suppressed. Figures 10(c)–10(f) illustrate the control effect for suppressing the acoustic cluster filtering outputs, $\tilde{p}_1$, $\tilde{p}_2$, $\tilde{p}_3$, and $\tilde{p}_4$, respectively. Compared with Fig. 6 which shows the case before control, it may be seen that the reductions obtained were 9 dB and 23 dB for the (1,1) and (1,3) modes in $\tilde{p}_1$, 20 dB and 13 dB for the (1,2) and (1,4) modes in $\tilde{p}_2$, 24 dB and 20 dB for the (2,1) and (2,3) modes in $\tilde{p}_3$, and 18 dB for the (2,2) mode in $\tilde{p}_4$.

V. CONCLUSION

Acoustic cluster control has been shown to be a successful approach to achieving global attenuation of structure-borne noise with a minimum computational load. It was shown that for control using a number of point sensors and point actuators, if the acoustic impedance matrix bridging the point actuators and point sensors is the normal matrix, then acoustic cluster control can be achieved. The validity of the acoustic cluster control was verified with a numerical simulation using four microphones and four speakers. By placing these sensors and actuators symmetrically with respect to the center of the gravity of the target panel, acoustic cluster filtering as well as cluster actuation was realized. Experimental results verified the numerical results for acoustic cluster filtering using four microphones and acoustic cluster actuation using four speakers. Using both acoustic cluster filtering and acoustic cluster actuation, acoustic cluster control was implemented experimentally and the capability of the approach for achieving global structure-borne noise suppression was demonstrated.

[1] C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (E&F Spon, London, 1997).
[2] O. L. Angevine, "Active cancellation of the hum of large electric transformers," *Proceedings of Inter Noise 92*, 3-3-316, Toronto, Canada (1992).
[3] O. L. Angevin, "Active control of hum from large power transformers: the real world," *Proceedings of the Second Conference on Recent Advances in Active Control of Sound and Vibration*, 279–290 (1993).
[4] O. L. Angevin, "The prediction of transformer noise," Sound Vib. Oct., 16–18 (1995).
[5] O. L. Angevin, "Active systems for attenuation of noise," Int. J. Active Control **1**, 65–78 (1995).
[6] M. J. Balas, "Active control of flexible systems," J. Optim. Theory Appl. **25**(3), 415–436 (1978).
[7] M. J. Balas, "Direct velocity feedback control of large space structure," J. Guidance **2–3**, 252–253 (1979).
[8] J. N. Aubrun, "Theory of the control of structures by low-authority control," J. Guid. Control **3**, 444–451 (1980).
[9] N. Tanaka and S. D. Snyder, "Cluster control of a distributed-parameter planar structure—Middle authority control," J. Acoust. Soc. Am. **112**(6), 2798–2807 (2002).
[10] R. Bellman, *Introduction to Matrix Analysis* (McGraw–Hill, New York, 1970).
[11] H. F. Harmuth, *Transmission of Information by Orthogonal Function*, 2nd ed. (Springer, Berlin, 1972).
[12] R. J. Pinnington and D. C. R. Pearce, "Multipole expansion of the vibration transmission between a source and a receiver," J. Sound Vib. **142**, 461–479 (1990).
[13] S. D. Snyder and N. Tanaka, "To absorb or not absorb: Control source power output in active noise control systems," J. Acoust. Soc. Am. **94**(1), 185–195 (1993).
[14] N. Tanaka, S. D. Snyder, and C. H. Hansen, "Distributed parameter modal filtering using smart sensors," ASME J. Vibr. Acoust. **118**, 630–640 (1996).
[15] N. C. Burgan, S. D. Snyder, N. Tanaka, and A. C. Zander, "A general approach to modal filtering for active noise control—Part I: iubration sensing," IEEE Sens. J. **2**(6), 577–589 (2002).

Numerical modeling of the sound fields in urban squares

Jian Kang[a)]
School of Architecture, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

(Received 26 June 2004; revised 29 December 2004; accepted 21 March 2005)

This paper studies the basic characteristics of sound fields in urban squares surrounded by reflecting building façades and the effectiveness of architectural changes and urban design options. A radiosity model and an image source model are developed, and a parametric study is carried out in hypothetical squares. The results show that the reverberation time (RT) is rather even in a square, whereas the early decay time (EDT) is very low in the near field, and then becomes close to RT after a rapid increase. Compared to diffuse boundaries, with geometrical boundaries the RT and EDT are significantly longer and the sound pressure level (SPL) attenuation with distance is generally smaller unless the height/side ratio is high. With a boundary diffusion coefficient of 0.2, the sound field is already close to that resulting from purely diffusely reflecting boundaries. The SPL in far field is typically 6–9 dB lower if the square side is doubled; 8 dB lower if the height of building façades is decreased from 50 m to 6 m (diffuse boundaries); 5 dB (diffuse boundaries) or 2 dB (geometrical boundaries) lower if the length/width ratio is increased from 1 to 4; and 10–12 dB lower if the boundary absorption coefficient is increased from 0.1 to 0.9. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1904483]

PACS numbers: 43.50.Rq, 43.50.Sr, 43.28.Fp [DKW] Pages: 3695–3706

I. INTRODUCTION

Squares are important elements of the urban environment, and their acoustic conditions are receiving great attention, especially with increased city center regeneration. However, most existing research in urban acoustics has concentrated on the noise reduction for buildings, whereas studies on the acoustic conditions of urban open public spaces, including squares, has been more scarce. Investigations mostly concentrate on a subjective evaluation of the acoustic environment.[1–3] In this research the sound propagation and soundscape in urban squares are systematically studied, including computer simulation of the sound fields, field measurement, questionnaire survey, semantic differential analysis of the soundscape, and an overall system for soundscape description, evaluation and design.[3–8]

The aim of this paper is to study basic characteristics of sound fields in urban squares surrounded by reflecting building façades and the effects of architectural changes and urban design options, including boundary reflection pattern, square size, height of building façades, aspect ratio, boundary absorption, and building and absorption arrangements. Two computer models have been developed: one based on the radiosity method, for diffusely reflecting boundaries; and the other based on the image source method, for geometrically reflecting boundaries. A beam-tracing program has also been used. The paper starts with a brief introduction of the two computer models; it then presents results of a parametric study.

II. SIMULATION MODELS

The two models presented below are for two kinds of idealized boundary condition, diffusely and geometrically reflective, respectively. The models are energy based. In other words, wave effects are ignored and the results are frequency independent. Since square dimensions are normally considerably greater than the wavelengths of urban sound sources, this is not problematic.

A. Radiosity model

The radiosity method was first developed for the study of radiant heat transfer in simple configurations,[9] and the techniques have also been used in computer graphics and lighting simulation,[10–12] as well as in room acoustics simulation for diffusely reflecting boundaries.[13–18] An important feature of using the radiosity method in acoustics is that the time factor should be considered for reverberation which can increase the computation time significantly compared with computer graphics and lighting simulations.

Using the radiosity techniques, the author has developed computer models for simulating sound propagation in urban streets.[19–22] A comparison between calculations using the radiosity method and measurements in a scale model of a typical street with Haussmann facades has shown a good agreement, within an accuracy of ± 1.5 dB for the sound pressure level (SPL), and 6% for the reverberation time (RT).[21] The algorithms have also been shown to correctly calculate the acoustic characteristics of a cube,[16] which can be regarded as a further validation. In this research, a model for urban squares has been developed. Since the basic principles and algorithms are similar to those of the street model, which have been presented elsewhere,[19–22] only a brief description of the square model is given below.

The model divides the square ground and building façades into a number of small ''patches'' with each patch represented by a node in a network, where receivers are also included. The sound propagation is simulated using the energy exchange between the patches. The energy exchange

a)Tel: +44 114 222 0325; Fax: +44 114 279 8276; electronic mail: j.kang@sheffield.ac.uk

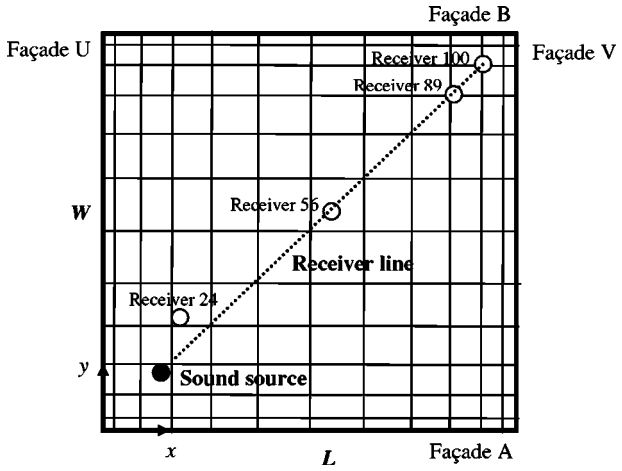


FIG. 1. Plan view of a typical urban square used in the computation, showing the division of patches, as well as source and receiver positions. There are 100 evenly distributed receiver positions in the square (which are different from the patch nodes shown in the figure). For the 50×50 m square, the source is at (10 m, 10 m), and the positions of four typical receivers are 24 (12.5 m, 17.5 m), 56 (27.5 m, 27.5 m), 89 (42.5 m, 42.5 m), and 100 (47.5 m, 47.5 m), corresponding to source–receiver distances of 8 m, 25 m, 46 m, and 53 m, respectively. The receiver line is from (12.5 m, 12.5 m) to (47.5 m, 47.5 m), with 8 receivers on the line.

between pairs of patches depends on a form factor, which is the fraction of sound energy emitted from one patch which arrives at the other by direct energy transport. The model assumes that all the boundaries are diffusely reflective according to the Lambert cosine law.

Consider an idealized rectangular square, as shown in Fig. 1. The first step is to divide each of the boundaries, including four facades A, B, U, V and ground G, into patches. The boundaries are so divided that a patch is smaller when it is closer to an edge since if the size was constant the form factor calculation would become less accurate. For convenience, in the model the patches are all rectangular, and the patch size varies in the manner of a geometric series. There is a precalculation process of form factors, where an optimal patch division can be determined in terms of calculation time and accuracy.

In a square as illustrated in Fig. 1, all the form factors are nonzero since there is always a line of sight between any pair of patches. There are two kinds of relative location between patches, either parallel or orthogonal. For the former, the form factor is calculated by projecting the receiving patch onto the upper half of a cube centered about the radiation patch. For the latter, computing a form factor is equivalent to projecting the receiving patch onto a unit hemisphere centered about the radiation patch, then projecting this projected area orthographically down onto the hemispere's unit circle base, and finally dividing by the area of the circle. In both cases, the boundary absorption coefficient is considered in the form factor calculation.

The energy exchange includes three steps: energy distributed from an impulsive source to the patches, energy exchange between patches, and energy radiation from patches to receivers.[19] The energy fraction from an omni-directional point source to a patch can be determined by the ratio of the solid angle subtended by the patch at the source to the total solid angle. Each patch is then regarded as an energy source, which is expressed in a form of energy response. If the source is directional, the energy fraction in each direction should be adjusted accordingly. The form factors are independent of the source position, and the energy exchange process between patches depends only on the form factors and the patch energy after the preceding energy exchange. This ''memoryless'' feature can significantly reduce the requirement for computer storage. The calculation of a kth order patch source is made by summing the contribution from all the $(k-1)$th order patch sources, except those which are on the same boundary as the kth order patch source considered. As an example, the contribution of the $(k-1)$th order patch sources on façade A to a kth order patch source on the ground, $G_k(t)_{l,m}$, can be calculated by

$$G_{Ak}(t)_{l,m}=\sum_{l'=1}^{N_X}\sum_{n'=1}^{N_Z} AG_{(l',n'),(l,m)}A_{k-1}\left(t-\frac{d_{(l',n'),(l,m)}}{c}\right)_{l',n'}\left(t-\frac{d_{(l',n'),(l,m)}}{c}\geqslant 0\right), \tag{1}$$

where $AG_{(l',n'),(l,m)}$ is the form factor, namely energy emitted from patch $A_{(l',n')}$ $(l'=1-N_X,\ n'=1-N_Z)$ to patch $G_{(l,m)}$, $A_{k-1}(t-d_{(l',n'),(l,m)}/c)_{l',n'}$ is the sound energy on the patch source $A_{l',n'}$ after the $(k-1)$th energy exchanges, and N_X and N_Z are patch numbers along x and z directions.

While energy is traveling between patches, the contribution of each patch to each receiver is calculated. By taking the direct sound transfer from source to receiver into account, an energy response can then be obtained for each receiver, from which the steady-state SPL, RT, and early decay time (EDT), which have been proved to be correlated to urban square acoustic comfort evaluation,[3,4,7,8] can be determined. The energy exchange process stops when the total energy on all patches reduces to a certain amount. For reverberation calculation this is typically 10^{-6} of the source energy, but if only SPL is concerned, this value could be much greater, so that the calculation time is shorter.

B. Image source model

The image source method has been commonly used to calculate the sound propagation in indoor and outdoor spaces with geometrically reflecting boundaries.[23–27] Consider the idealized rectangular square as shown in Fig. 1. By treating the building facades and square ground as mirrors, a series of image sources can be created, as illustrated in Fig. 2. With a given image source, the reflected sound is modeled with a sound path directly from the image source to a receiver. The energy response at a receiver is determined by summing the contribution from all the image sources, and taking the direct sound transfer into account. The acoustic indices, including steady-state SPL, RT, and EDT, can consequently be determined.

For the convenience of calculating the distance to receiver R, the image sources in Fig. 2 are divided into a

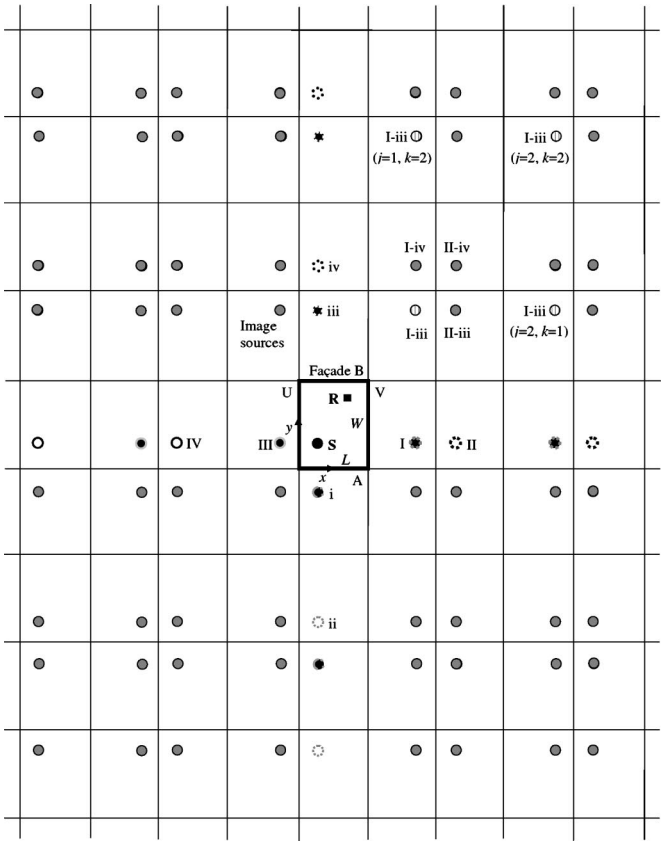


FIG. 2. Plan view of a rectangular square, showing the distribution of image sources.

number of groups, such as I-iii, II-iii, I-iv, and II-iv, where I, II, III, and IV correspond to four basic distances in the x direction, and i, ii, iii, and iv correspond to four basic distances in the y direction. Consider an image source (j,k) $(j=1\cdots\infty;\ k=1\cdots\infty)$ in group I-iii, for example, the energy from an image source to a receiver R at (R_x,R_y,R_z) can be determined by

$$E_{j,k}(t)=\frac{1}{4\pi d_{j,k}^2}(1-\alpha_A)^{k-1}(1-\alpha_B)^k(1-\alpha_U)^{j-1}\times(1-\alpha_V)^j e^{-Md_{j,k}}\quad\left(t=\frac{d_{j,k}}{c}\right),\tag{2}$$

where α_A, α_B, α_U, and α_V are the absorption coefficient of façades A, B, U, and V, respectively. M (Np/m) is the intensity related attenuation constant in air. $d_{j,k}$ is the distance from the image source (j,k) to the receiver,

$$d_{j,k}^2=(2jL-S_x-R_x)^2+(2kW-S_y-R_y)^2+(S_z-R_z)^2.\tag{3}$$

By considering the ground reflection, an image source plane similar to that in Fig. 2 can be obtained. The energy from the image sources to receiver R can be determined in a similar manner as above, but replacing term S_z-R_z with S_z+R_z and also taking the ground absorption α_G into account.

III. COMPUTATION

These two models are applied in this paper to carry out a parametric study with a series of hypothetical urban squares based on a recent survey.[7,8] The radiosity method is used for diffusely reflecting boundaries and the image source method for geometrically reflecting boundaries. While the models are accurate for simulating the two kinds of idealized sound field, which is the main attention of this paper, it is also useful to consider the conditions in between, namely with partially diffuse and partially geometrical boundaries, or in other words, with a diffusion coefficient $0<d<1$. For this purpose Raynoise,[28] a commercial acoustic simulation software based on beam-tracing, is used.

First a typical configuration is considered, analyzing basic characteristics of sound fields in urban squares and comparing the sound fields formed by the two kinds of boundary. Then the effects of square size, height of building façades, square shape, boundary absorption, and building and absorption arrangements are analyzed. Indices for comparison include the steady-state SPL, RT (RT30 in this paper) and EDT.

A. Sound field in a typical square

The analysis in this section is based on a square of 50 $\times$50 m. The square is surrounded by buildings with a height of 20 m. Since the buildings are assumed to have a constant height in a square, a term ''square height'' is used in this paper for convenience. The absorption coefficient of all the boundaries is 0.1. A point source, with a sound power level of 0 dB, is positioned at ($x=10$ m, $y=10$ m, $z=1.5$ m). 100 receivers are evenly distributed in the square, at a height of 1.2 m.

Figure 3(a) shows the SPL distribution in the square, with both diffusely and geometrically reflecting boundaries. It can be seen that with both kinds of boundary, the SPL initially decreases significantly with increasing source-receiver distance, about 5–8 dB from 8 m to 25 m (corresponding to receiver 24–56 as shown in Fig. 1), for example, and then becomes approximately stable, with a variation of less than 2 dB beyond 25 m. This feature is similar to that in regularly-shaped enclosures, although the SPL values are different from those calculated using the classical theory.

The RT and EDT are rather long in such a square, about 2 s with diffusely reflecting boundaries, and around 8–10 s with geometrically reflecting boundaries, as shown in Figs. 3(b) and (c), respectively. It can also be seen that the RT is very even over the entire square. The standard deviation to average ratio is 0.9% for diffusely reflecting boundaries and 3.3% for geometrically reflecting boundaries. The EDT is very low in the near field, say 10–15 m from the source, and then becomes relatively even after a rapid increase. Note in this paper the terms ''near field'' and ''far field'' refer to close and remote receiver positions, rather than dimensions relating to the size and wavelengths of the sound source. The standard deviation to average ratio is 28.8% for diffusely reflecting boundaries and 32.5% for geometrically reflecting boundaries. When the source–receiver distance is increased from 8 m to 25 m, changes in RT and EDT are plus 10%–

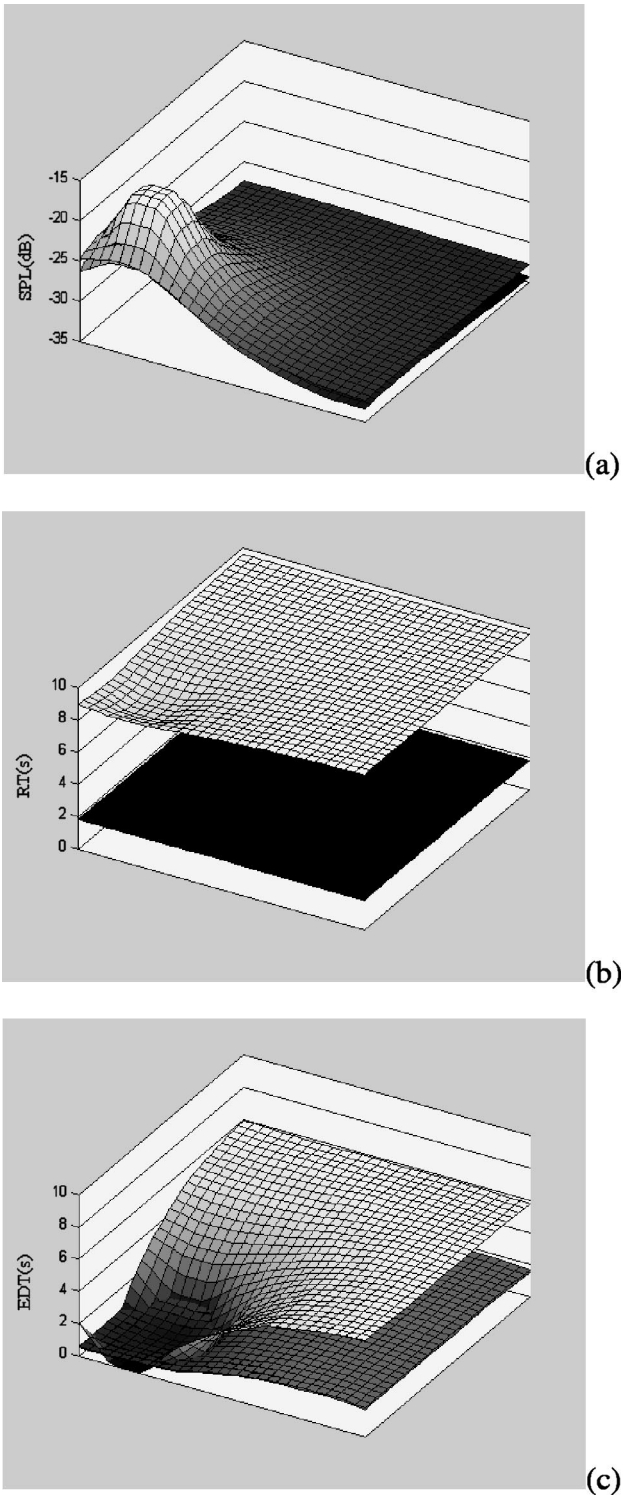


FIG. 3. Distribution of the steady-state SPL (a), RT (b), and EDT (c) in a square of 50×50 m, with diffusely (dark mesh) and geometrically (light mesh) reflecting boundaries. Square height is 20 m. Source at (10 m, 10 m). Boundary absorption coefficient 0.1.

20% and plus 70%–110%, respectively, with both kinds of boundary considered. Beyond 25 m the variations are much less, <3% and 2%–8% for RT and EDT. An important reason for the short reverberation in the near field, especially the EDT, is that the direct sound plays an important role, and there is a lack of early reflections.

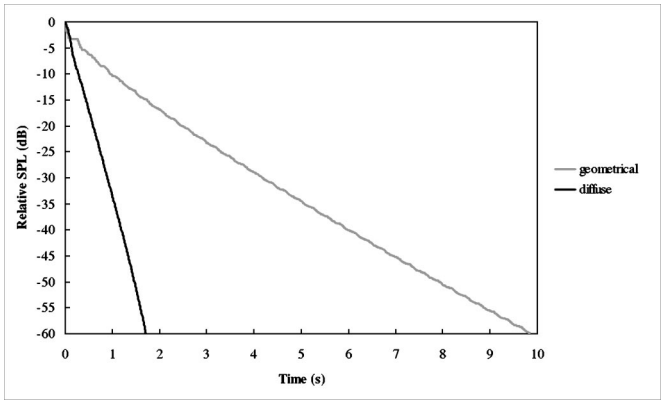


FIG. 4. Comparison of the decay curves between diffusely and geometrically reflecting boundaries at receiver 100, where the source–receiver distance is 53 m, as shown in Fig. 1. Square is 50×50 m. Square height 20 m. Source at (10 m, 10 m). Boundary absorption coefficient 0.1.

B. Boundary reflection pattern

In urban streets and irregularly-shaped enclosures it has been demonstrated that there are considerable differences in sound fields resulting from diffusely and geometrically reflecting boundaries.[15–22] In this section a comparison between the two kinds of boundary is made for urban squares.

From Fig. 3 it can be seen that the SPL with diffusely reflecting boundaries is generally lower than that with geometrically reflecting boundaries. This difference increases with increasing source–receiver distance, to a maximum of about 2 dB. The main reason for the difference is that with diffusely reflecting boundaries some energy is redistributed towards the source due to backscattering. A 2 dB difference is considerable given that for both kinds of boundary the variation in SPL across the receivers, except in the very near field, is only about 5–8 dB. More explicit comparisons between the two kinds of boundary can be seen in the following sections, where receivers along a diagonal are considered.

The reverberation resulting from geometrically reflecting boundaries is significantly longer at about 400% for the RT and 200% for the EDT, than that from diffusely reflecting boundaries. An important reason for the differences is that for a given order of reflection, with geometrically reflecting boundaries the sound path is generally much longer, mainly due to the flutter echo effect. This can be seen in Fig. 4, where the decay curves with diffusely and geometrically reflecting boundaries are compared at receiver 100. The main reason for the significant difference in reverberation between the two kinds of boundary compared to that in SPL is that the former depends on multiple reflections, whereas the latter is mainly determined by early reflections. Similar results have also been observed in street canyons.[19]

Ray-tracing type models,[24–26] although less accurate compared to the image source method and the radiosity method for geometrically and diffusely reflecting boundaries respectively, are useful for studying the effect of boundary diffusion coefficient. Using Raynoise, variations in SPL, RT, and EDT with increasing d ranging from 0 to 1 are calculated, as shown in Fig. 5. Three receivers, 24, 56, and 89, representing typical distances from the source, as illustrated

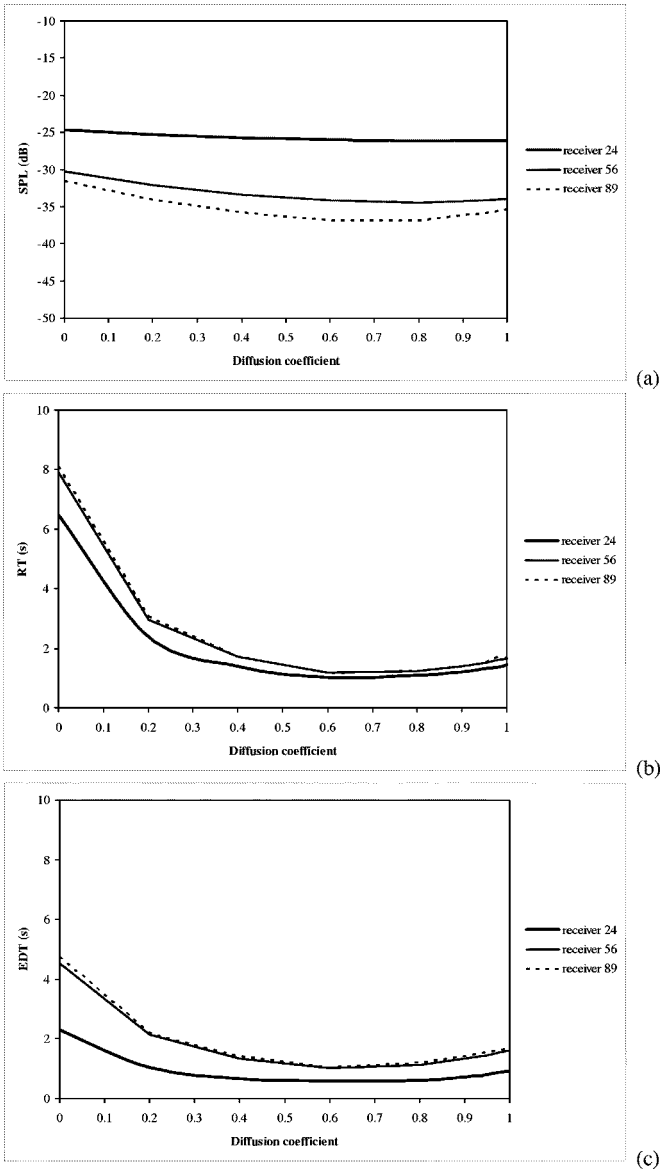


FIG. 5. Variation of the steady-state SPL, RT, and EDT with increasing diffusion coefficient, calculated using Raynoise, at receiver 24, 56, and 89, where the source–receiver distances are 8 m, 25 m, and 46 m, respectively, as shown in Fig. 1. Square is 50×50 m. Square height 20 m. Source at (10 m, 10 m). Boundary absorption coefficient 0.1.

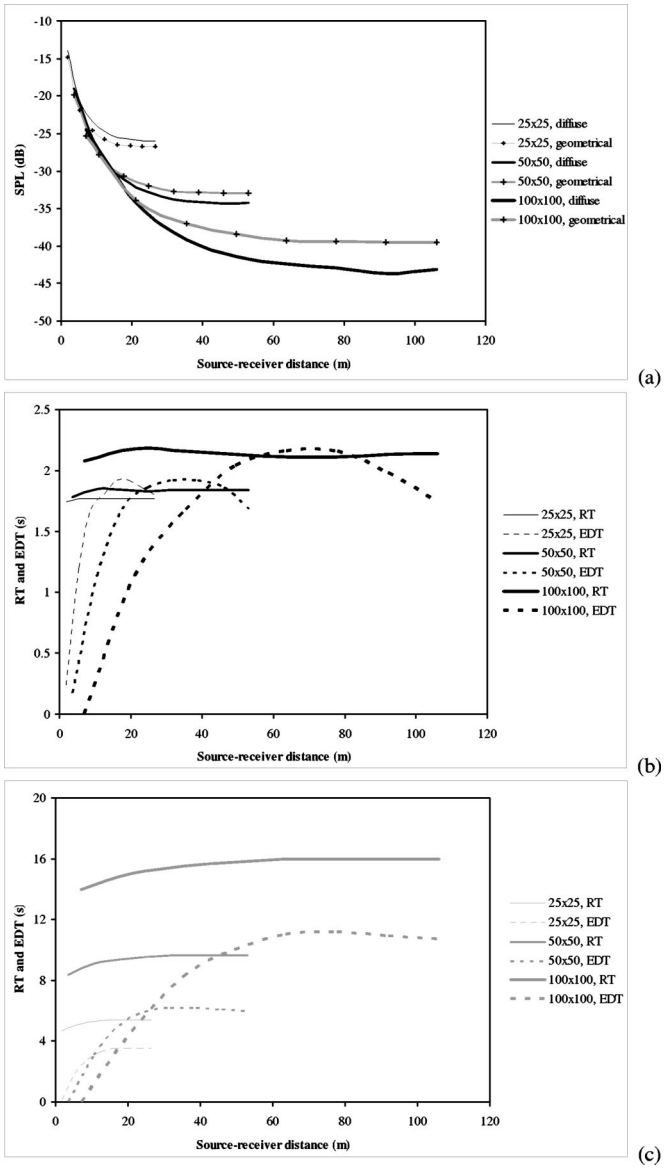


FIG. 6. Comparison between three square sizes: 25×25 m, 50×50 m, and 100×100 m. (a) Steady-state SPL with both diffusely and geometrically reflecting boundaries; (b) RT and EDT with diffusely reflecting boundaries; (c) RT and EDT with geometrically reflecting boundaries. Square height is 20 m. Source at (5 m, 5 m), (10 m, 10 m), and (20 m, 20 m) in the three squares, respectively. Receivers along the diagonal as illustrated in Fig. 1. Boundary absorption coefficient is 0.1.

in Fig. 1, are considered. In the calculation a large ray number, 200 k, and a high reflection order, 100, are used. Since random numbers are involved in the Raynoise computation when a diffusion coefficient is considered, causing variations in results, the data in Fig. 5 are based on an average of several computations. From Fig. 5 it can be seen that with partially diffuse and partially geometrical boundaries, the sound field is generally between the two idealized situations.

It is interesting to note that when d is increased from 0 to about 0.2, the decrease in RT and EDT is significant, whereas when d is further increased, the changes become much less. The variation in SPL is similar to that in RT and EDT, although the d value where the SPL becomes approximately steady is slightly greater than 0.2. This means that with only about 20% of the energy incident upon the boundaries diffusely reflective, the sound field in an urban square is close to that resulting from purely diffusely reflecting boundaries. Similar results have been obtained in other spaces, and the importance of considering boundary diffusion has been discussed by other researchers.[29–33] In many urban squares in Europe it might be appropriate to use the sound field resulting from purely diffusely reflecting boundaries as an approximation of the actual sound field, since the façades and ground normally have a certain level of diffusion. In addition, various kinds of square ''furniture'' are useful for diffusing sound.

It is also noted in Fig. 5 that when d is increased from about 0.7 to 1, there is a slight increase in RT and EDT, which is probably caused by the decrease in initial energy in the sound decay process. With increasing source–receiver distance, the variation in SPL, EDT, and RT with various d

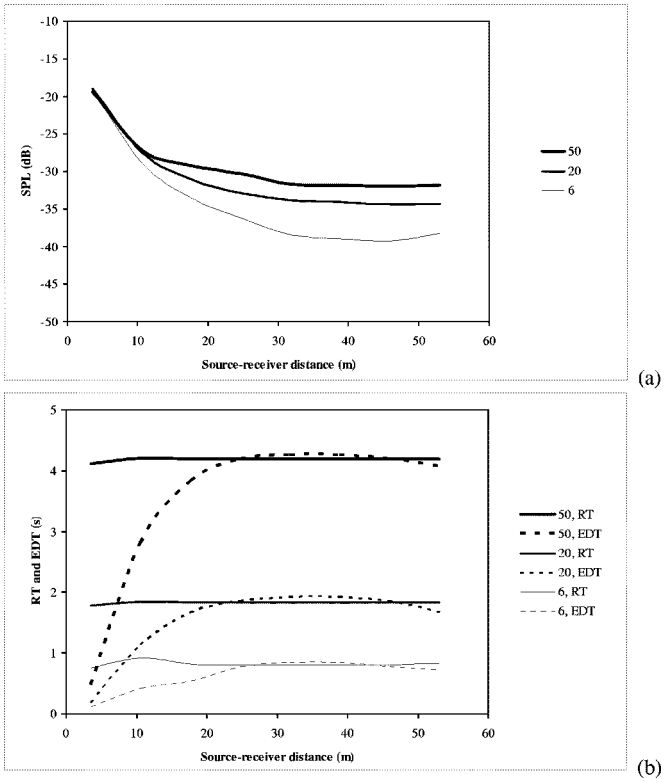


FIG. 7. Comparison between three square heights: 50, 20, and 6 m. The boundaries are diffusely reflective. (a) Steady-state SPL; (b) RT and EDT. Square is 50×50 m. Source at (10 m, 10 m). Receivers along the diagonal as illustrated in Fig. 1. Boundary absorption coefficient is 0.1.

generally becomes slightly greater, as can be seen by comparing the three receivers in Fig. 5. This is possibly because reflections play a more important role at longer distances to the source.

C. Size

The SPL distributions in three squares of different sizes, 25×25 m, 50×50 m, and 100×100 m, are compared in Fig. 6(a), where the square height is 20 m and the boundary absorption coefficient is 0.1. Both diffusely and geometrically reflecting boundaries are considered. The receivers are along a diagonal of each square, as illustrated in Fig. 1.

In the near field, say within a source–receiver distance of 5–10 m, there is no significant difference in SPL between the three square sizes. Clearly this is due to the dominant role of the direct sound. With the increase of source–receiver distance, the SPL becomes systematically less with increased square size. In the far field of each square, the SPL is approximately 6–9 dB lower when the square side is doubled.

Compared to geometrically reflecting boundaries, with diffusely reflecting boundaries there are more energy losses resulting from the scattering of sound out of the square. Moreover, the SPL attenuation with distance is greater since some energy is redistributed towards the source due to backscattering, as mentioned in the previous section. However, with diffusely reflecting boundaries, reflections from all boundary patches will contribute to the SPL at a receiver, whereas with geometrically reflecting boundaries only the surfaces below the source and the receiver heights are effective. The importance of these three effects varies with different square configurations. In Fig. 6(a), with the square sizes of 50×50 m and 100×100 m, the first two effects are dominant, so that with diffusely reflecting boundaries the SPL attenuation is systematically greater than that with geometrically reflecting boundaries, at about 2–5 dB. This corresponds to the results in long enclosures, urban streets and flat enclosures.[15,30] With the increase of height/side ratio, the third effect plays a major role and thus with diffusely reflecting boundaries, the SPL attenuation is systematically smaller than that with geometrically reflecting boundaries. This is the case of the 25×25 m square in Fig. 6(a), although the difference is within 1 dB.

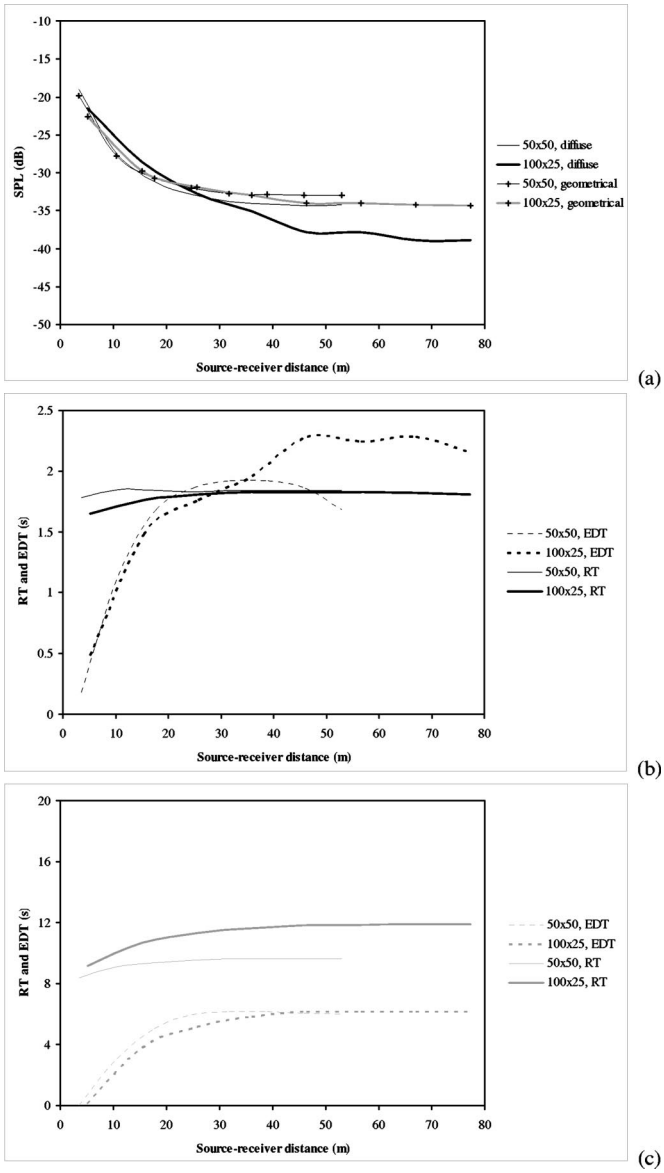
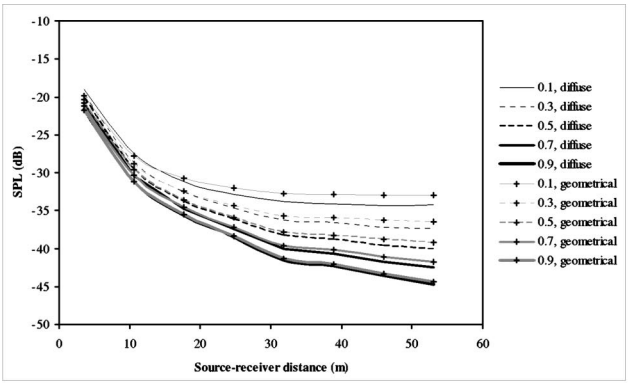
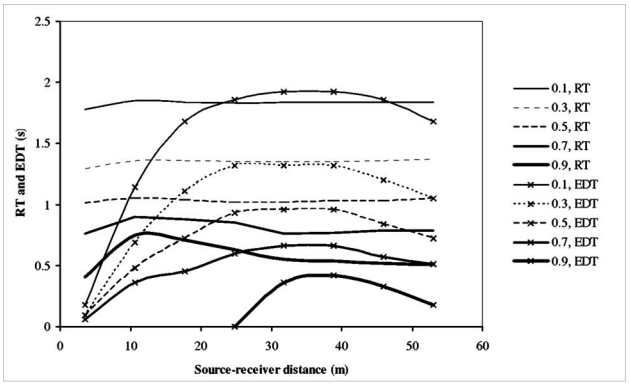


FIG. 8. Comparison between two square shapes: 50×50 m and 100×25 m. (a) Steady-state SPL with both diffusely and geometrically reflecting boundaries; (b) RT and EDT with diffusely reflecting boundaries; (c) RT and EDT with geometrically reflecting boundaries. Square height is 20 m. Source at (10 m, 10 m) and (20 m, 5 m) in the two squares, respectively. Receivers along the diagonal similar to that illustrated in Fig. 1. Boundary absorption coefficient is 0.1.

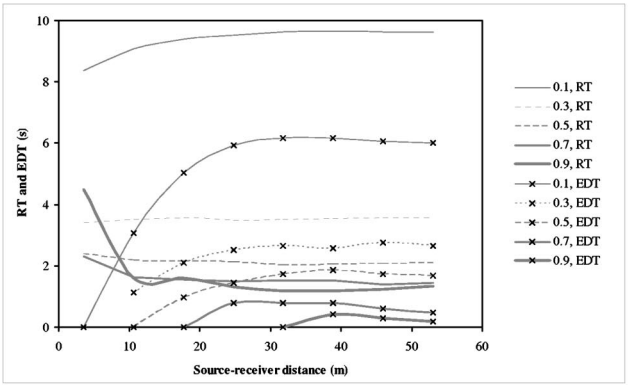
The reverberation results with diffusely reflecting boundaries are shown in Fig. 6(b). The RT increases with



(a)



(b)



(c)

FIG. 9. Effects of boundary absorption in the 50×50 m square. (a) Steady-state SPL with both diffusely and geometrically reflecting boundaries; (b) RT and EDT with diffusely reflecting boundaries; (c) RT and EDT with geometrically reflecting boundaries. Square height is 20 m. Source at (10 m, 10 m). Receivers along the diagonal as illustrated in Fig. 1.

increasing square size, as expected. It is interesting to note that the simulated RT values in the three squares, 1.77 s, 1.84 s, and 2.14 s, are close to the calculations using the classical Eyring formula for room acoustics,[34] 1.88 s, 2 s, and 2.03 s, where the ceiling absorption coefficient is assumed as 1. However, using the two methods the RT ratios between various square sizes, namely, 1.77:1.84:2.14 and 1.88:2:2.03, are different, suggesting that the direct use of the Eyring formula may not be appropriate. In terms of the distribution of RT and EDT, the three squares are similar: the RT is even across a square, whereas the EDT increases rapidly with increasing source–receiver distance until about the square center, and then becomes relatively stable, where the EDT is close to RT, indicating that the decay curves are close to linear. It is noted that the increase in EDT with increasing source–receiver distance is slower in a larger square. This is mainly because the sound paths become longer with increased square size, so that there is a lack of initial reflections.

With geometrically reflecting boundaries, the RT and EDT variations are generally similar to the above, as shown in Fig. 6(c), although the absolute values as well as the RT and EDT ratios between various square sizes are rather different. Unlike diffusely reflecting boundaries, with geometrically reflecting boundaries the RT values are considerably higher than the EDT. This is probably due to the large number of late reflections corresponding to flutter echoes.

D. Height

The effect of square height on the SPL distribution is shown in Fig. 7(a), where the square size is 50×50 m and the boundary absorption coefficient is 0.1. Three square heights are calculated, 50 m, 20 m, and 6 m. Only diffusely reflecting boundaries are considered. With geometrically reflecting boundaries the acoustic indices at a receiver including SPL, RT, and EDT depend only on the boundaries below the source or the receiver height, whichever is greater, and thus, increasing square height from 6 m to 50 m will have no effect on receivers at 1.5 m. From Fig. 7(a) it can be seen that in the near field the SPL is almost unchanged with various square heights, whereas with the increase of source-receiver distance, the SPL becomes higher with a greater square height. This is expected, because with more façade areas there are more reflections. With increasing source–receiver distance the SPL difference between different square heights initially increases, and then becomes approximately stable. In the far field, between square heights 50 m and 6 m, the difference is about 8 dB.

The RT and EDT with the three square heights are compared in Fig. 7(b), again for diffusely reflecting boundaries only. The simulated RT values, averaged over the entire square, are 4.19 s, 1.84 s, and 0.83 s with square heights of 50 m, 20 m, and 6 m, which are close to the calculated values using the Eyring formula, 4.31 s, 2 s, and 0.63 s. In all the three squares the EDT becomes stable beyond about 25 m from the source, although for a greater square height the increase in EDT with increasing source–receiver distance is more rapid, because of the increased reflections by the increased boundary area. In the relatively far field, the EDT values are close to RT with various square heights, indicating that the decay curves are close to linear.

E. Shape

It has been demonstrated that with a constant floor area, sound distributions in long, flat and regularly-shaped enclosures differ considerably.[17] In Fig. 8(a) the SPL attenuation in two squares with the same ground area but different aspect ratios, 50×50 m ($L/W=1$) and 100×25 m ($L/W=4$), are compared, where the square height is again 20 m and the boundary absorption coefficient is 0.1. It is interesting to note that for both diffusely and geometrically reflecting boundaries, with $L/W=4$, the SPL in the near field is almost the same as, or only slightly higher than, that with $L/W=1$, whereas in the far field, say over 30 m, the SPL with $L/W=4$ is considerably lower, especially with diffusely re-

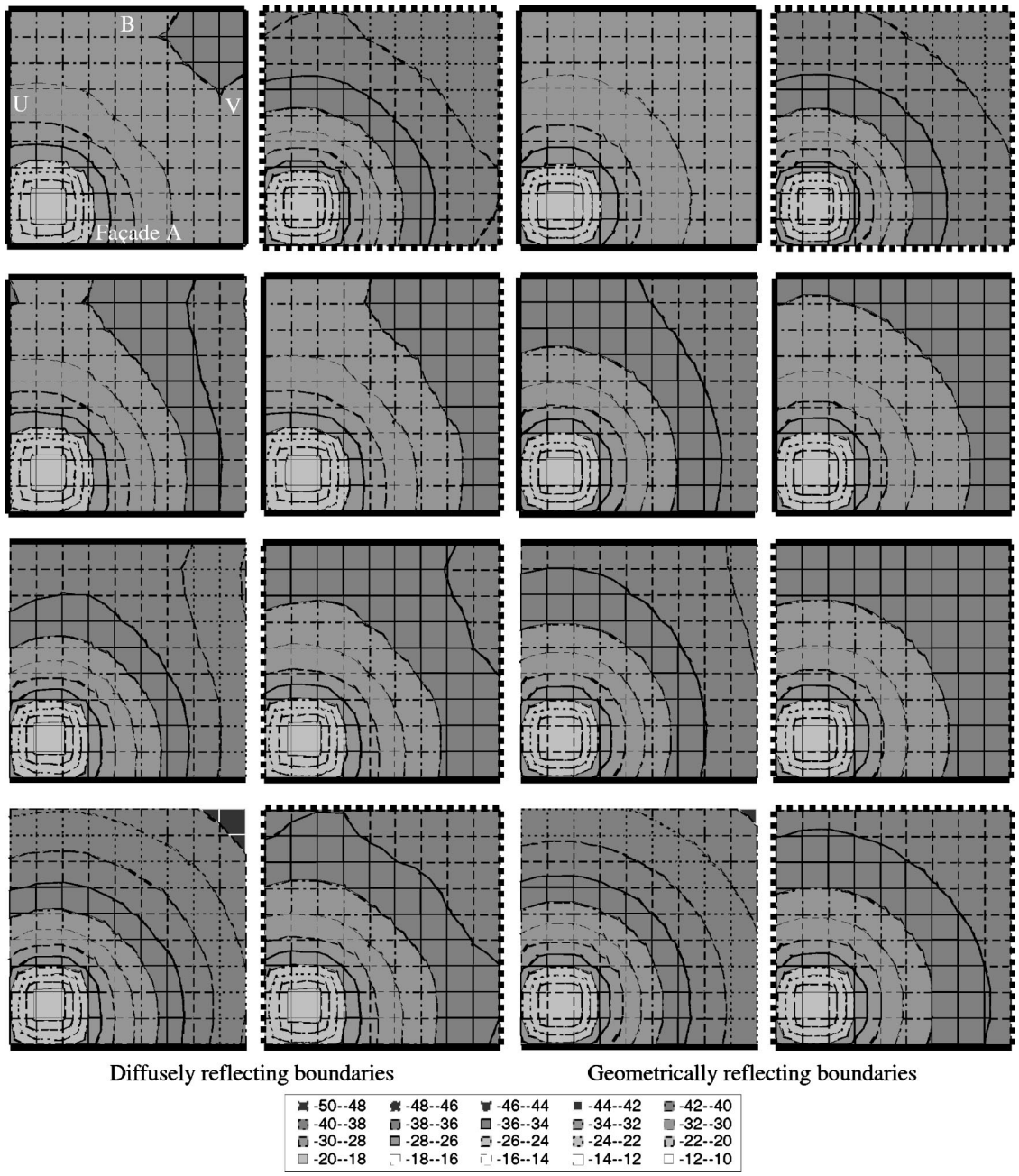


FIG. 10. Steady-state SPL distribution with various building and absorption arrangements in the 50×50 m square. Solid lines, $\alpha=0.1$; dotted lines, $\alpha=0.5$; no line, $\alpha=1$ (e.g., no building). Square height is 20 m. Source at (10 m, 10 m).

flecting boundaries. A main reason for the extra SPL attenuation is that the sound path is generally greater when a square becomes longer. The results suggest that from the viewpoint of noise reduction, it is better to design a square with a greater aspect ratio.

It is noted that the SPL differences between diffusely and geometrically reflecting boundaries becomes greater when the aspect ratio of a square increases. As can be seen in Fig. 8(a), for the 50×50 m square, the difference between the two kinds of boundary is about 2 dB in the far field, whereas this value is about 5 dB for the 100×25 m square. A possible reason is that with diffusely reflecting boundaries, the relative increase in sound path when a square becomes longer is generally greater than that with geometrically reflecting boundaries. Also, when using diffusely reflecting boundaries, a longer square will backscatter more energy to the near field, as can be seen in Fig. 8(a).

The differences in reverberation between the two square shapes are generally less than those in SPL, both for diffusely and geometrically reflecting boundaries, as shown in Figs. 8(b) and (c), respectively. This is probably because the SPL is more affected by early reflections and reflection patterns, whereas the reverberation is more related to multiple reflections and the square volume. In Fig. 8(b) it is noted that with diffusely reflecting boundaries, in the far field of the 100×25 m square the EDT is considerably longer than RT. Clearly this is caused by the strong early reflections relative to the direct sound in this area. With geometrically reflecting boundaries the RT in the 100×25 m square is longer than that in the 50×50 m square, by about 18% on average, and

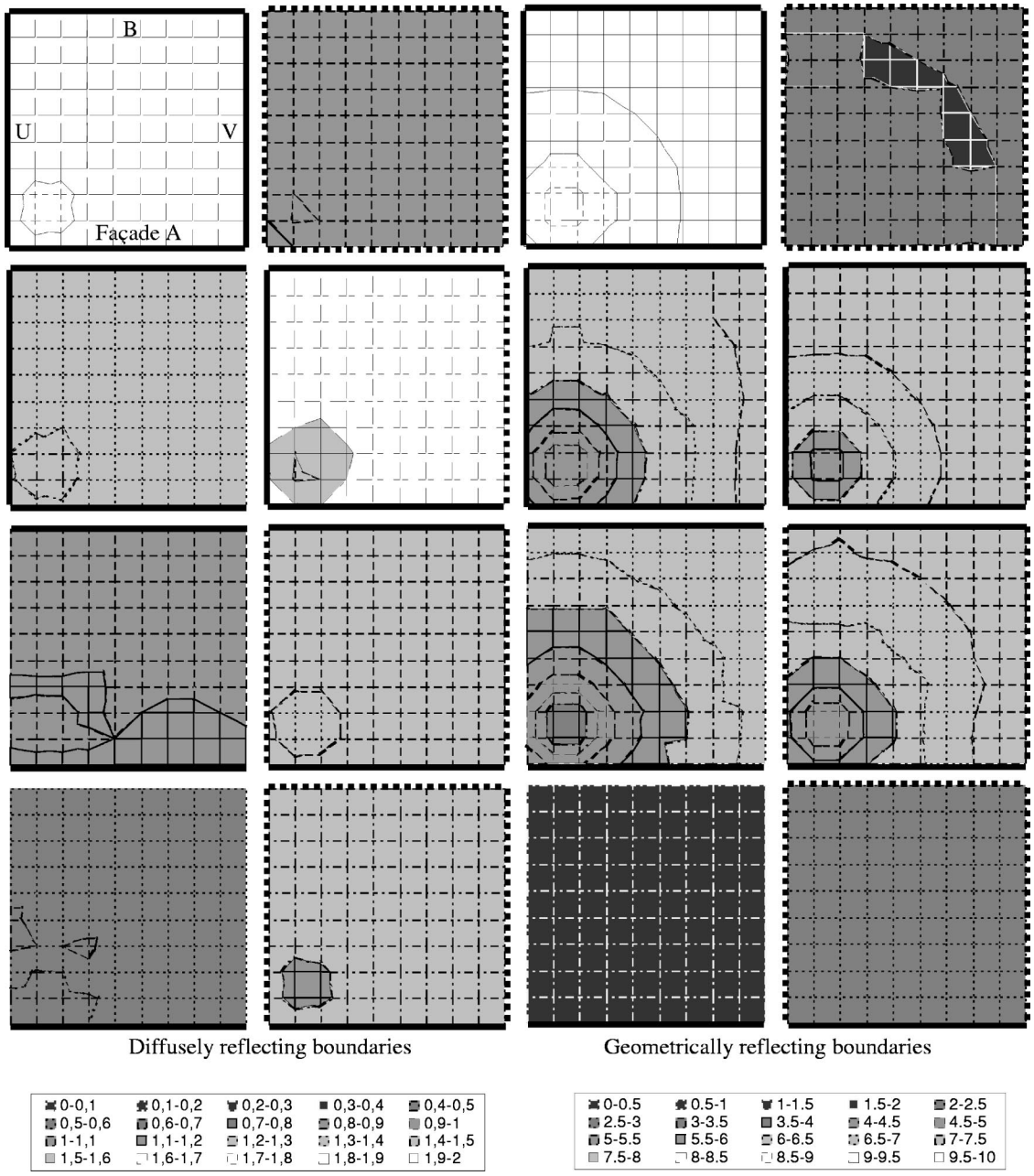


FIG. 11. RT distribution with various building and absorption arrangements in the 50×50 m square. Solid lines, $\alpha=0.1$; dotted lines, $\alpha=0.5$; no line, $\alpha=1$ (e.g., no building). Square height is 20 m. Source at (10 m, 10 m).

this difference increases with increasing source–receiver distance, as can be seen in Fig. 8(c). This is probably because in the 100×25 m square the sound paths are generally longer and thus in the decay process the late reflections are relatively stronger compared to the 50×50 m square. For a similar reason, the near field EDT in the 100×25 m square is slightly shorter than that in the 50×50 m square, due to the lack of early reflections.

F. Boundary absorption

Figure 9(a) shows the SPL attenuation with a range of boundary absorption coefficients, where the square is 50 ×50 m and the square height is 20 m. When the absorption coefficient is increased from 0.1 to 0.3, 0.3 to 0.5, 0.5 to 0.7, and 0.7 to 0.9, the SPL decrease averaged over the entire square is 2.0 dB, 1.7 dB, 1.5 dB, and 1.3 dB, respectively, for diffusely reflecting boundaries, and 2.4 dB, 1.6 dB, 1.4 dB, and 1.3 dB, respectively, for geometrically reflecting boundaries. This indicates that a given increase of sound absorption is less efficient for noise reduction if there are already considerable absorbers in the square, which is expected. It is also noted in Fig. 9(a) that the decrease in SPL caused by increasing absorption coefficient becomes systematically greater with increasing source–receiver distance. Clearly this is because with a longer source–receiver distance the SPL depends more on reflections, for which boundary absorption is effective.

In Fig. 9(a) the maximum SPL difference between absorption coefficients 0.1 and 0.9 is 10.4 dB with diffusely reflecting boundaries and 11.5 dB with geometrically reflecting boundaries. These values are considerably greater than

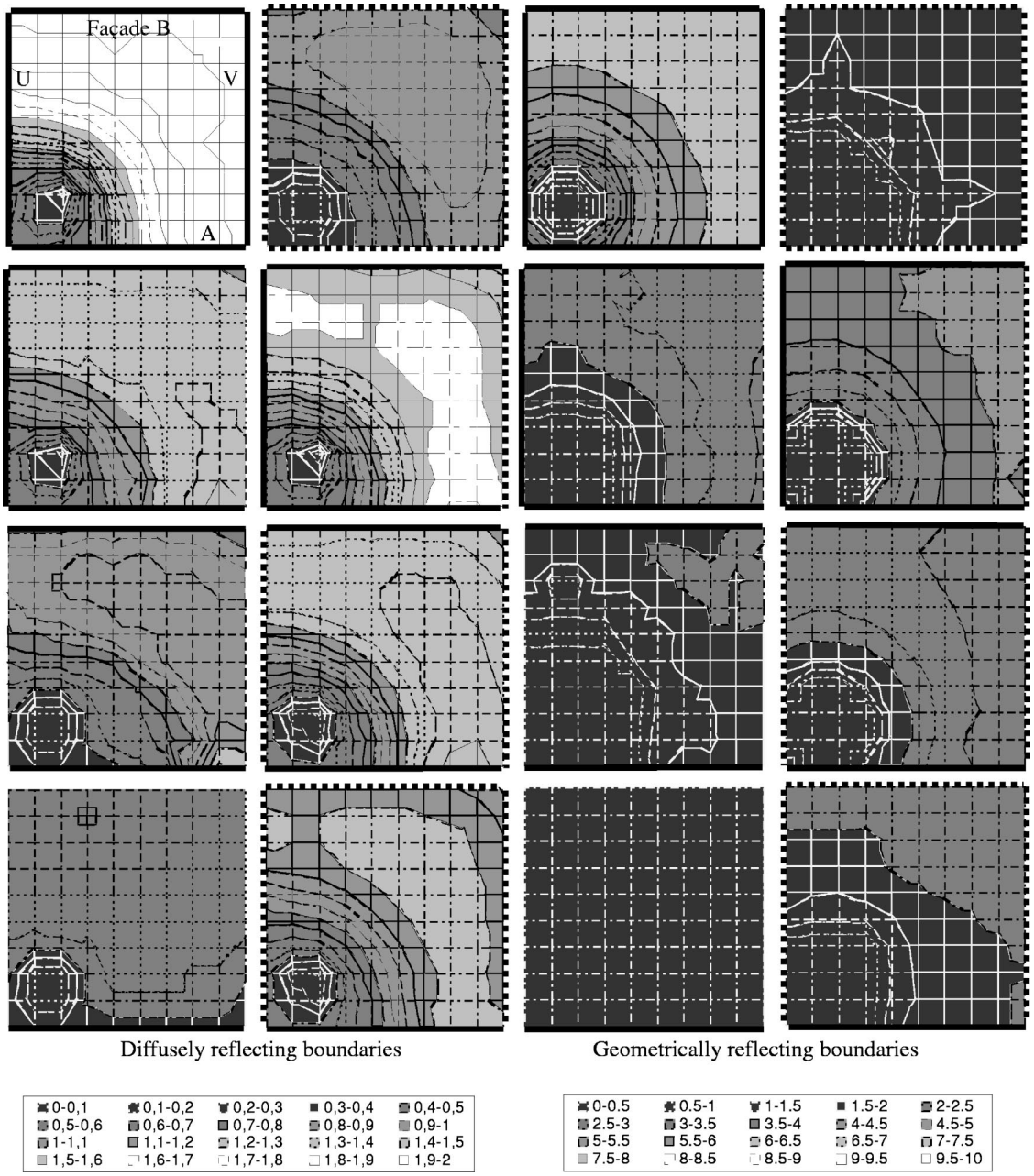


FIG. 12. EDT distribution with various building and absorption arrangements in the 50×50 m square. Solid lines, $\alpha=0.1$; dotted lines, $\alpha=0.5$; no line, $\alpha=1$ (e.g., no building). Square height is 20 m. Source at (10 m, 10 m).

that in street canyons.[19–21] The main reason is that in urban squares the SPL depends on reflections from all four facades and thus the effect of increasing boundary absorption is more significant. It is also seen in Fig. 9(a) that the difference between the two kinds of boundary becomes less with increasing absorption coefficient. This is because with a higher absorption coefficient, boundary reflections become less important compared to the direct sound.

The RT decreases with increasing boundary absorption, as expected, but it is interesting to note that the rate of decrease is considerably slower than that predicted using the Eyring formula. This is true for both kinds of boundary, as can be seen in Figs. 9(b) and (c), respectively. This suggests the nondiffuse feature of the sound fields in urban squares, especially when the boundary absorption is high. For both diffusely and geometrically reflecting boundaries, the RT/EDT ratio generally becomes greater with increasing boundary absorption. This is probably because a given SPL reduction in reflections can cause a greater decrease in EDT than that in RT.

When the boundary absorption coefficient is high, say 0.7–0.9, the energy in early reflections becomes very weak, especially in the near field where the number of reflections is small. The shapes of the decay curves thus become extremely nonlinear, especially for geometrically reflecting boundaries. As a result, the RT and EDT vary considerably at various receivers, as shown in Figs. 9(b) and (c).

G. Building and absorption arrangement

Urban squares are not always surrounded by buildings and sound absorption is often unevenly distributed. To study

such effects, simulation is carried out using the 50×50 m square, with one, two and three facades only, as well as with absorbers of $\alpha=0.5$ on one to four facades. The other boundaries have an absorption coefficient of 0.1. Both diffusely and geometrically reflecting boundaries are considered.

The SPL results are shown in Fig. 10. The tendencies with both kinds of boundary are generally similar. When facade V is removed or made absorbent, the SPL near this façade becomes lower, showing a rather ''local'' effect. When two opposite facades, U and V, are removed or made absorbent, the direct sound plays a much more important role, although the effect of multiple reflections between façades A and B can still be seen. When three façades, U, V, and B, are removed, the sound field is dominated by the direct sound. When the three facades are absorbent, the SPL distribution shows a similar pattern.

The RT distribution with various configurations is shown in Fig. 11. With diffusely reflecting boundaries, the RT is generally even across the square with various building and absorption arrangements, although the RT value reduces with decreased number of façades or increased number of absorbent facades, as expected. The standard deviation to average ratios are 0.9%, 1.4%, 3.2%, and 5.9% with four, three, two, and one facade, and 1.5%, 1.6%, 1.9%, and 2.2% with one, two, three, and four façades absorbent. With geometrically reflecting boundaries the reflection pattern in the square is significantly changed with different building and absorption arrangements and consequently, the RT distribution becomes much more uneven. The standard deviation to average ratios are 3.3%, 11.4%, and 12.7% with four, three, and two facades, and 4.7%, 7.4%, 9.2%, and 1.6% with one, two, three, and four façades absorbent.

The effects of building and absorption arrangements on the EDT are shown in Fig. 12. The tendencies of changes are generally similar to those of RT, but the variations are much greater. For various configurations the standard deviation to average ratio ranges 25%–34% with diffusely reflecting boundaries and 32%–110% with geometrically reflecting boundaries.

IV. CONCLUSIONS

The sound fields in urban squares have been analyzed and compared through computations in a number of hypothetical squares using the radiosity and image source models. It has been shown that: (1) With either diffusely or geometrically reflecting boundaries, the SPL initially decreases significantly with increasing source–receiver distance and then becomes approximately stable. The RT and EDT are rather long, about 2 s with diffusely reflecting boundaries and around 8–10 s with geometrically reflecting boundaries in a typical square of 50×50 m. The RT is very even across a square, whereas the EDT is very low in the near field, and then becomes close to RT after a rapid increase. (2) Compared to diffusely reflecting boundaries, with geometrically reflecting boundaries the RT and EDT are significantly longer, typically by 200%–400%, and the SPL attenuation along a square is generally smaller unless the height/side ratio is high, say 1:1. (3) When the boundary diffusion coefficient is increased from 0 to about 0.2, the decrease in SPL, RT, and EDT is significant, whereas when d is further increased, the changes become much less. (4) If a relatively far field is considered, the SPL is typically 6–9 dB lower when the square side is doubled; 8 dB lower when the square height is decreased from 50 m to 6 m (diffusely reflecting boundaries); 5 dB (diffusely reflecting boundaries) or 2 dB (geometrically reflecting boundaries) lower if the length/width ratio is increased from 1 to 4; and 10–12 dB lower if the boundary absorption coefficient is increased from 0.1 to 0.9. (5) The Eyring formula can give an approximate indication of the changes in RT and EDT with varied square size, height, and shape, but the rate of RT and EDT decrease caused by increasing boundary absorption is considerably slower than that predicted by the Eyring formula. (6) When one facade is removed or made absorbent, the sound field near this facade is mostly affected, whereas when two or more facades are removed or made absorbent, the direct sound plays a much more important role. With diffusely reflecting boundaries the RT is generally even across a square with various building or absorption arrangements, whereas with geometrically reflecting boundaries the RT distribution becomes much more uneven.

ACKNOWLEDGMENTS

The author is indebted to Dr. M. Zhang, W. Yang, and Dr. M. Brocklesby for useful discussions. This research was a part of Key Action 4 ''City of Tomorrow and Cultural Heritage'' from the program ''Energy, Environment and Sustainable Development'' within the Fifth Framework Program of the European Commission.

[1] B. Berglund, C. A. Eriksen, and M. E. Nilsson, ''Perceptual characterization of soundscapes in residential areas,'' in Proceedings of the 17th International Conference on Acoustics (ICA), Rome, Italy (2001).
[2] M. Raimbault, M. Bérengier, and D. Dubois, ''Common factors in the identification of urban soundscapes pilot study in two French cities: Lyon and Nantes,'' in Proceedings of the 17th International Conference on Acoustics (ICA), Rome, Italy (2001).
[3] J. Kang and W. Yang, ''Soundscape in urban open public spaces,'' World Archit. **144**, 76–79 (2002) (in Chinese).
[4] J. Kang and M. Zhang, ''Semantic differential analysis on the soundscape of urban open public spaces,'' J. Acoust. Soc. Am. **112**, 2435 (2002).
[5] J. Kang, ''Computer simulation of the sound fields in urban squares: comparison between diffusely and geometrically reflecting boundaries,'' in Proceedings of the 32nd International Acoustical Conference (IAC)—European Acoustics Association (EAA) Symposium, Slovakia (2002), pp. 157–160.
[6] J. Kang and M. Zhang, ''Acoustic simulation and soundscape in urban squares,'' in Proceedings of the 10th International Congress on Sound and Vibration, Stockholm, Sweden (2003), pp. 1405–1412.
[7] W. Yang and J. Kang, ''Acoustic comfort evaluation in urban open public spaces,'' Appl. Acoust. **66**, 211–229 (2005).
[8] W. Yang and J. Kang, ''Soundscape and sound preferences in urban squares,'' Journal of Urban Design **10**, 61–80 (2005).
[9] R. Siegel and J. Howell, *Thermal Radiation Heat Transfer*, 2nd ed. (Hemisphere, Washington, D.C., 1981).
[10] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principle and Practice*, 2nd ed. (Addison–Wesley, Reading, 1990).
[11] F. X. Sillion and C. Puech, *Radiosity and Global Illumination* (Morgan Kaufmann, New York, 1994).
[12] A. Tsangrassoulis, M. Santamouris, D. Asimakopoulos, and P. R. Tregenza, ''A method for the estimation of illuminances on surfaces of urban canyons with balconies in sunlit areas,'' Light. Res. Technol. **31**, 5–12 (1999).

[13] G. R. Moore, ''An approach to the analysis of sound in auditoria,'' Ph.D. dissertation, University of Cambridge, UK, 1984.
[14] T. Lewers, ''A combined beam tracing and radiant exchange computer model of room acoustics,'' Appl. Acoust. **38**, 161–178 (1993).
[15] J. Kang, *Acoustics of Long Spaces: Theory and Design Practice* (Thomas Telford, London, 2002).
[16] J. Kang, ''Reverberation in rectangular long enclosures with diffusely reflecting boundaries,'' Acta Acust. Acust. **88**, 77–87 (2002).
[17] J. Kang, ''Comparison of sound fields in regularly-shaped, long and flat enclosures with diffusely reflecting boundaries,'' Int. J. Acoust. Vib. **7**, 165–171 (2002).
[18] J. Kang, ''Numerical modelling of the speech intelligibility in dining spaces,'' Appl. Acoust. **63**, 1315–1333 (2002).
[19] J. Kang, ''Sound propagation in street canyons: Comparison between diffusely and geometrically reflecting boundaries,'' J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[20] J. Kang, ''Sound propagation in interconnected urban streets: a parametric study,'' Environ. Plan. B: Plan. Des. **28**, 281–294 (2001).
[21] J. Kang, ''Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries,'' J. Sound Vib. **258**, 793–813 (2002).
[22] J. Kang, Y. Meng, and G. Brown, ''Sound propagation in micro-scale urban areas: simulation and animation,'' in Proceedings of the 5th European Conference on Noise Control (Euronoise), Naples, Italy (2003).
[23] J. B. Allen and D. A. Berkley, ''Image method for efficiently simulating small-room acoustics,'' J. Acoust. Soc. Am. **65**, 943–950 (1979).
[24] J. Borish, ''Extension of the image model to arbitrary polyhedra,'' J. Acoust. Soc. Am. **75**, 1827–1836 (1984).
[25] M. Vorländer, ''Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm,'' J. Acoust. Soc. Am. **86**, 172–178 (1989).
[26] G. M. Naylor, ''ODEON—Another hybrid room acoustical model,'' Appl. Acoust. **38**, 131–143 (1993).
[27] D. J. Oldham and M. M. Radwan, ''Sound propagation in city streets,'' Build. Acoust. **1**, 65–88 (1994).
[28] LMS Numerical Technologies, *Raynoise Users' Manual: Building Acoustics and Industrial Noise Simulation* (LMS, Leuven, Belgium, 2001).
[29] R. H. Lyon, ''Role of multiple reflections and reverberation in urban noise propagation,'' J. Acoust. Soc. Am. **55**, 493–503 (1974).
[30] M. R. Hodgson, ''Evidence of diffuse surface reflections in rooms,'' J. Acoust. Soc. Am. **89**, 765–771 (1991).
[31] J. Picaut, L. Simon, and J. Hardy, ''Sound field modelling in a street with a diffusion equation,'' J. Acoust. Soc. Am. **106**, 2638–2645 (1999).
[32] J. Picaut, ''Numerical modeling of urban sound fields by a diffusion process,'' Appl. Acoust. **63**, 965–991 (2002).
[33] M. R. Ismail and D. J. Oldham, ''Computer modelling of urban noise propagation,'' Build. Acoust. **10**, 227–259 (2003).
[34] H. Kuttruff, *Room Acoustics*, 3rd ed. (Applied Science, London, 1989).

Evaluation of decay times in coupled spaces: Reliability analysis of Bayeisan decay time estimation

Ning Xiang[a)]
School of Architecture and Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180

Paul M. Goggans
Department of Electrical Engineering, University of Mississippi, University, Mississippi 38677

Tomsilav Jasa and Mendel Kleiner
School of Architecture and Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180

(Received 27 January 2005; revised 6 March 2005; accepted 7 March 2005)

This paper discusses quantitative tools to evaluate the reliability of ''decay time estimates'' and inter-relationships between multiple decay times for estimates made within a Bayesian framework. Previous works [Xiang and Goggans, J. Acoust. Soc. Am. **110**, 1415–1424 (2001); **113**, 2685–2697 (2003)] have applied Bayesian framework to cope with the demanding tasks in estimating multiple decay times from Schroeder decay functions measured in acoustically coupled spaces. A parametric model of Schroeder decay function [Xiang, J. Acoust. Soc. Am. **98**, 2112–2121 (1995)] has been used for the Bayesian model-based analysis. The relevance of this work is that architectural acousticians need to know how well determined are the estimated decay times calculated within Bayesian framework using Schroeder decay function data. This paper will first address the estimation of global variance of the residual errors between the Schroeder function data and its model. Moreover, this paper discusses how the ''landscape'' shape of the posterior probability density function over the decay parameter space influences the individual decay time estimates, their associated variances, and their inter-relationships. This paper uses experimental results from measured room impulse responses in real halls to describe a model-based sampling method for an efficient estimation of decay times, and their individual variances. These parameters along with decay times are relevant decay parameters for evaluation and understanding of acoustically coupled spaces. [DOI: 10.1121/1.1903845]

PACS numbers: 43.55.Br, 43.55.Mc [EJS] Pages: 3707–3715

I. INTRODUCTION

In recent years, acoustics of coupled spaces have received considerable attention in architectural acoustics.[1–7] One of the important tasks for better understanding and design of acoustics in coupled spaces lies in sound energy decay analysis. This paper discusses quantitative tools to evaluate the reliability of ''decay time estimates'' and inter-relationships between multiple decay times for estimates made within a Bayesian framework. Recent applications of Bayesian probability theory in decay time evaluation in acoustically coupled spaces[4,5] have demonstrated a useful framework for analyzing Schroeder decay functions[8] from room impulse response measurements. The Bayesian framework proves to be able to estimate not only the decay parameters from Schroeder decay model,[4] but also to determine the decay order quantitatively.[5]

Bayesian probability theory, which includes Bayes' theorem, includes all rules of probability theory for relating and manipulating probabilities; interpreted as *the logic of science*,[9] it is a quantitative theory of inference. In Bayesian framework, Bayesian probability theory provides useful tools to formulate posterior probability density function (PPDF) of decay parameters. One approach to estimation of respective decay parameters is to use the PPDF in probabilistic moment calculations of specific parameters. Alternatively, localization of global extremes of the PPDF in the parameter space will also lead to an effective estimation of relevant decay parameters,[4] so-called maximum *a posterior* (MAP) estimation, since the PPDF associated with any particular set of values for the decay parameters is a measure of how much one believes that they really lie in the neighborhood of that range. For both probabilistic approaches, the shape of the global extremes on the PPDF will influence the precision of the probabilistic estimation, subsequently the reliability of the decay time estimation. In addition to estimation of decay times in order to understand the acoustics in the acoustically coupled spaces, recent investigation on systematic dependence of the decay times on the aperture size, the volumes and other room parameters[6,7] stimulates the need to evaluate reliability, to quantify uncertainties associated with Bayesian decay time estimation from the room impulse responses.

Reliability of measurements and estimation of parameters of interest is always of major concern of the respective measurement techniques and estimation methods. In the con-

[a)]Author to whom correspondence should be addressed. Electronic mail: xiangn@rpi.edu; Tel: (518)-276-6464; Fax: (518)-276-3034.

text of architectural-acoustic measurement techniques, acousticians have always endeavored a high reliability in determining room-acoustic parameters including reverberation time, decay times, and other parameters.[8–13] Schroeder backwards integration method[8] is a successful effort to increase the determination reliability of sound energy decay function from a single measurement of room impulse responses, in comparison to the traditional noise-interrupt method, where a large number of ensemble averages of measured decay curves have to be carried out.

The subject of this paper is to study the estimation reliability and inter-relationships between multiple decay times given the measured room impulse responses from which Schroeder decay functions are derived, in other words, this is to access the estimation quality of relevant decay parameters from Schroeder decay functions using Bayesian probability theory. A practical significance of the reliability analysis is that architectural acousticians want to know in the practice what error bars or variances are associated with the estimated decay parameters once the decay parameters have been estimated from the given data. Reliability associated with different measurement techniques, associated with reproducibility of a specific measurement technique, or different estimation methods is, however, not a primary concern of the current work, rather than the reliability of the estimated decay parameters when the Bayesian inference is applied in analyzing Schroeder decay functions.[4] This is however a different situation than assessing result-spreading, error-bars or the reproducibility of a number of repeated measurements since Schroeder integration yields a smoothed decay function from a single measurement of a room impulse response.[8]

Fundamentals of Bayesian reliability analysis can be found in Refs. 9 and 14. Bayesian uncertainty analysis has also been applied in other acoustics fields, amongst others, in previous works.[15–18] In this paper, however, we share our own insight in the specific applications of sound energy decay analysis in coupled spaces. We will demonstrate that Bayesian framework also provides architectural acousticians with relevant tools for quantifying the reliability of decay time estimation and inter-relationships between multiple decay times. This paper will demonstrate the reliability analysis of decay time estimation in coupled spaces using experimentally measured room impulse responses.

This paper is organized as follows: Sec. II provides some examples of calculating decay times in acoustic spaces in order to highlight estimation reliability of decay times over their PPDF landscapes. Section III introduces the formulations to quantify the decay time reliability within Bayesian framework and discusses some representative results. Finally Sec. IV concludes the paper.

II. DECAY TIME LANDSCAPES

Bayesian decay time estimation is a model-based approach. Previous work[4] described the Bayesian formalism using a parametric model[19] for Schroeder decay function

$$F(\mathbf{A},\mathbf{B},t_k)=\sum_{j=1}^{m} A_j G_j(B_j,t_k),\quad 0\leqslant t_k\leqslant t_{\mathrm{ULI}}, \tag{1}$$

where

$$G_j(B_j,t_k)=\begin{cases}\exp(-B_j\cdot t_k) & \text{for } j=1,2,...,m-1,\\ t_{\mathrm{ULI}}-t_k & \text{for } j=m,\end{cases} \tag{2}$$

and $\mathbf{A}=\{A_1,A_2,\ldots,A_m\}$, $\mathbf{B}=\{B_1,B_2,\ldots,B_m\}$, and $B_j=13.8/T_j$ and $B_m=0$. t_{ULI} is the time at the upper limit of the Schroeder integration (ULI).[4,19] Here T_j denotes the jth decay time to be determined, and m is the number of additive terms in the model. The A_j are called linear parameters while B_j are nonlinear parameters. Since Schroeder integration of experimentally measured room impulse responses is carried out only within a finite ULI, one more boundary condition has to constrain the Schroeder decay model when the integration is carried out towards the ULI:

$$F(\mathbf{A},\mathbf{B},t_k=t_{\mathrm{ULI}})=0. \tag{3}$$

This boundary condition introduces an additional constant term into the model[20]

$$G_j(B_j,t_k)=\begin{cases}\exp(-B_j\cdot t_k)-\exp(-B_j\cdot t_{\mathrm{ULI}}) & \text{for } j=1,2,...,m-1,\\ t_{\mathrm{ULI}}-t_k & \text{for } j=m.\end{cases} \tag{4}$$

Simulation and experimental comparisons reveal that Eq. (2) differs from Eq. (4) only insignificantly when t_{ULI} is large enough so that the decay time analysis usually ignores the additional, constant term in the practice, especially when a measurement record length of room impulse responses using pseudorandom signals has to be selected long enough to avoid possible time-aliasing. A subsequent Schroeder integration will be simply carried out over the entire record-length without any burden of careful determination of the ULI.[4,19] However, when computer modeling tools create room impulse responses, there is no measurement noise in the data ($A_m=0$), the model in Eq. (4) will be more appropriate.

The model expressed in Eq. (1) along with Eq. (2) or Eq. (4) is a form of generalized linear models.[4] It is derived based on the nature of Schroeder decay functions with a few, distinct decay modes.[4,19,20] For this specific model form, Bayesian formulation can carry out marginalizations of linear parameters and other nuisance parameters analytically, leading to the PPDF,[4] which is only a function of the nonlinear parameters (decay times) given the Schroeder decay function. Figure 1 illustrates two examples of the PPDFs over the decay times based on the decay model when $m=3$ (two decay slopes). The sound energy decay analysis is primarily confined with common frequency bands in architectural acoustics. The Schroeder decay functions in Fig. 1

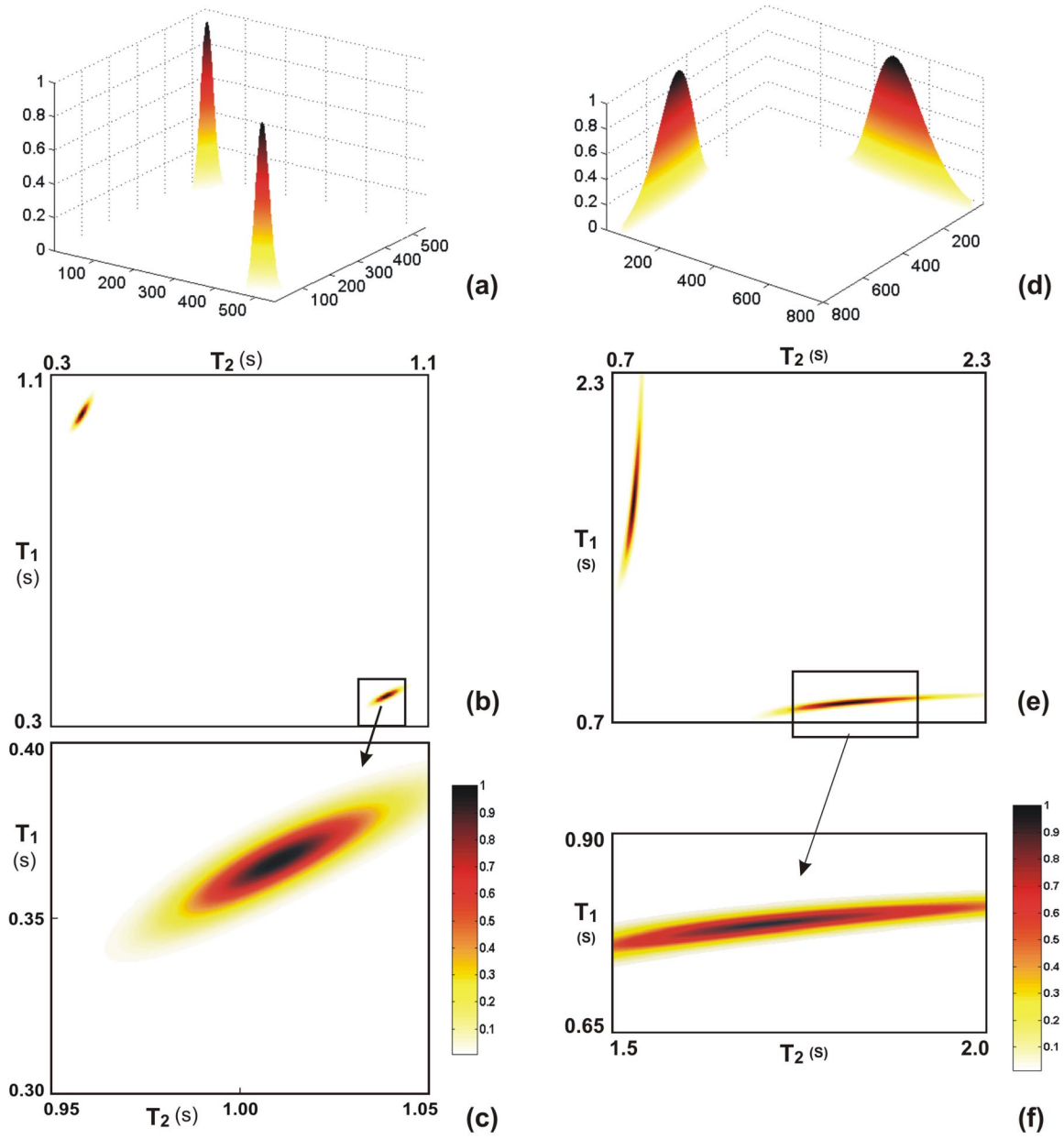


FIG. 1. Peak-normalized posterior probability distributions (PPDF) of two decay times evaluated from room impulse responses in two halls. Left: A Church in south Mississippi. Right: Coupled spaces in the Student Union, University of Mississippi. (a), (d): PPDF in 3D presentations, a grid of 550×550 and 800×800 is defined for (a), (d) respectively. (b), (e): PPDF in 2D presentations. (c), (f): Detailed PPDF showing elongated, skewed peaks in decay time spaces.

are e.g. evaluated from 1 kHz octave band-filtered, experimentally measured room impulse responses. Bayesian decay model selection[5] has determined the number of decay slopes to be 2. On the left-hand side of Fig. 1 the results, normalized with respect to the peak value (peak-normalized), are derived from a room impulse response measured in a church. Two sharp peaks of the PPDF are found within a decay time range between 0.3 and 1.1 s. Any of these peaks will serve when seeking the global maximum. A detailed zoom in one of the peaks as shown in Fig. 1(c) reveals a sharp, but skewed landscape of PPDF over the decay time space $\{T_1, T_2\}$. The sharpness of the landscape indicates a high reliability (low uncertainty) associated with the decay time estimation based on the PPDF. In other words, the decay times are well defined and will be precisely estimated within a tiny decay time range in an order of $|0.02|$ s. The skewness of the landscape, namely, the angle of the landscape relative to the axes, however, implies a dependence between two decay times.

On the right-hand side of Fig. 1 the results are derived from a room impulse response measured in the Student Union of the University of Mississippi where two spaces are coupled together. The PPDF landscape of this example is characterized by two distinct, well-separated, narrow peaks as shown in Figs. 1(d) and (e). A detailed zoom into one of the peaks as shown in Fig. 1(f) reveals an elongated landscape. The specific peak is widely spreading over a large range along decay time axis T_2, while the landscape along

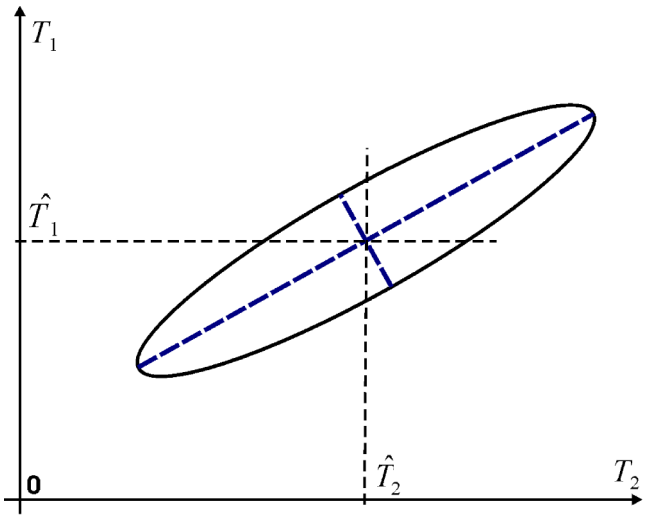


FIG. 2. Elliptical contour of the PPDF landscapes around the global peak over the decay time space. The length of the principle axes (dotted line) are related to the variance of each decay time and the orientation of the skewness corresponds to the dependence of one decay time on the other.

decay time axis T_1 presents a narrow peak. The elongated landscape indicates that decay time T_1 can be inferred more reliably than decay time T_2 on the basis of the current data. Estimation of T_1 will be associated with an order of magnitude higher reliability. The landscape oblique to the axis T_2 implies a dependence of one decay time from the other.

These two experimental examples illustrate the reliability associated with the decay time parameters and inter-relationship between them may vary from data to data. Once the decay model order has been determined[5] different measurement results feature different landscape shapes over the decay time space. Figure 2 shows an abstracted contour of the PPDF landscapes illustrated in Fig. 1, which indicates that an ellipse can approximate a contour of the PPDF landscape around the global peak at a given, sufficiently high PPDF value. Any point specified by $\{T_1, T_2\}$ on the ellipse will have equal value of the PPDF, which measures the same degree of our belief that the decay times to be determined are in the vicinity around that point. PPDF landscapes illustrated in these figures should clearly highlight different reliability associated with shapes of PPDFs and need of further evaluations of parameters associated with the PPDF landscapes, in addition to the Bayesian estimation of decay times[4] which yields only the linear decay parameters and decay times. Note that generations of the landscapes illustrated in Fig. 1 are quite computationally expensive, 550^2 and 800^2 calculations of the PPDF need to be undertaken, respectively. They are illustrated only for an clear explanation of the problem and our need of further evaluation. In architectural acoustics practice, efficient, quantitative tools are required to measure the reliability and inter-relationship between them without visual inspection of these landscapes. Section III B details the tools.

III. RELIABILITY ANALYSIS

In this section we discuss two kinds of quantitative measures of reliability: *global* and *individual* variance in the context of evaluation of multiple decay times from Schroeder's decay functions. This discussion is based heavily on the Bayesian decay time estimation detailed in Ref. 4 so as to keep the discussion as concise as possible.

A. Global variance

To formulate the PPDF of decay parameters given the data (Schroeder decay function),[4] application of the principle of maximum entropy results in a Gaussian probability density function for the residual errors between the Schroeder decay data d_k and its model $F(\mathbf{A},\mathbf{B},t_k)$ with an unspecified variance σ^2 [see Eq. (19) in Ref. 4] which is the likelihood function of data:

$$p(\mathbf{D}|\mathbf{A},\mathbf{B},\sigma,I)=(\sqrt{2\pi}\sigma)^{-K} \times \exp\left\{\frac{-1}{2\sigma^2}\sum_{k=1}^{K}[d_k-F(\mathbf{A},\mathbf{B},t_k)]^2\right\}, \quad (5)$$

where K is the total number of decay data used for estimation. The likelihood function is written $p(\mathbf{D}|\mathbf{A},\mathbf{B},\sigma,I)$ as a reminder that the variance σ^2 is at this stage unknown. The background information I encapsulating what is known about these parameters $(\mathbf{A},\mathbf{B},\sigma)$ before observing the data. It also specifies the degree we should believe the model reflecting a few distinct decay modes is correct for this specific problem.[4] Although the conditioning on I has been omitting in all the calculations presented in Refs. 4, 5 and this work, one must never forget its existence. "A failure to state explicitly all the relevant background information, and assumptions is frequently the real cause of heated debates about data analysis."[14]

In the following we make effort to estimate the variance σ^2. After marginalization of linear parameters $\mathbf{A}$ by assigning a uniform prior, the PPDF of nonlinear parameters becomes[4]

$$p(\mathbf{B}|\sigma,\mathbf{D},I)\propto\sigma^{-K+m}\exp\left(-\frac{K\overline{d^2}-m\overline{q^2}}{2\sigma^2}\right), \quad (6)$$

where m is the number of additive terms in the decay model, the quantities $\overline{d^2}$ and $\overline{q^2}$ are defined in Eqs. (6) and (20) in Ref. 4, respectively. $p(\mathbf{B}|\sigma,\mathbf{D},I)$ can be interpreted as a function of σ. Once the decay parameters $\mathbf{B}$ given the data vector $\mathbf{D}$ have been well estimated, the moment calculation by assigning Jeffreys' prior $1/\sigma$ (Ref. 21) yields the expected s-moment $\langle\sigma^s\rangle$,[22,23]

$$\langle\sigma^s\rangle=\frac{\int_0^\infty\sigma^{s-1}p(\mathbf{B}|\sigma,\mathbf{D},I)d\sigma}{\int_0^\infty\sigma^{-1}p(\mathbf{B}|\sigma,\mathbf{D},I)d\sigma}. \quad (7)$$

For $s=2$ it yields the expected variance (see Appendix):

$$\langle\sigma^2\rangle=\frac{K\overline{d^2}}{K-m-2}\left(1-\frac{m\overline{q^2}}{K\overline{d^2}}\right)_{|\mathbf{B}=\hat{\mathbf{B}}}, \quad (8)$$

where $\hat{\mathbf{B}}$ are estimated MAP values of nonlinear decay parameters.

The expected *variance* $\langle\sigma^2\rangle$ or its square root, the expected *standard derivation* $\langle\sigma\rangle$, serves as a global indicator

TABLE I. Overall standard derivation [Eq. (8)] of two different Schroeder's decay functions from experimentally measured impulse responses. "Union" stands for the Student Union at the University of Mississippi; "Church" for the Church in Pascagoula, Mississippi.

Hall	Overall standard derivation	T_1 (s)	T_2 (s)
Church	1.51E−3	0.374	1.02
Union	2.26E−3	0.77	1.59

of how well the specific decay model represents the data (Schroeder decay function), provided that the decay parameters have been well determined. The quantities K, m, $\overline{d^2}$, and $\overline{q^2}$ would have been determined when seeking the global maximum of the landscape at $\hat{\mathbf{B}}$ as discussed in Ref. 4 so that $\langle\sigma\rangle$ using Eq. (8) is readily evaluated. It is a quantitative indicator for quality assessment which quantifies the general variation of the measured decay function from the model function. For a given K and m, the smaller $\langle\sigma\rangle$, the smaller the variation will be. Table I lists the standard derivation σ evaluated from experimental measurements shown in Fig. 1.

Although the expected standard derivation $\langle\sigma\rangle$ can give us an idea how the Schroeder decay function is in general represented by the decay model when the decay parameters have been optimally estimated, it does not convey any specific information about reliability of each single decay times estimated from the data. Nevertheless, as a quantitative measure when the expected standard derivation $\langle\sigma\rangle$ becomes unexpected large, one needs to check with what went wrong with the measurement or the model. Since decay times in acoustically coupled spaces are of primary interest in architectural acoustics practice, we proceed to specify direct tools to quantify the reliability associated with each single decay time.

B. Individual variance

As stated previously, no specific knowledge prior to the decay analysis lead to assignment of uniform priors for all the parameters to be determined.[4] The data likelihood function has been assigned a Gaussian function due to limited knowledge on the residual errors between the data and the model. In the end, the likelihood distribution will place most of its weight, as a relatively sharp peak, in a region where the prior distribution varies little. In the current work, a large number of data are involved in calculating the likelihood function, since Schroeder decay functions are presented in data on orders of hundreds and thousands data points. An n-dimension Gaussian model will asymptotically approximate the PPDF landscape within a small subspace around a global peak[14,24] as illustrated in Figs. 1(c) and (f),

$$p(\mathbf{T}) \propto \exp[-\tfrac{1}{2}(\mathbf{T}-\hat{\mathbf{T}})\mathbf{C}^{-1}(\mathbf{T}-\hat{\mathbf{T}})^{\mathrm{Tr}}], \tag{9}$$

where $\mathbf{T}=\{T_1,T_2,\ldots,T_n\}$ are the decay times and $\hat{\mathbf{T}}=\{\hat{T}_1,\hat{T}_2,\ldots,\hat{T}_n\}$ the decay times at the global maximum of the landscape, also termed MAP-values of the decay times. $\mathbf{C}$ is the covariance matrix. $(\,)^{\mathrm{Tr}}$ represents the matrix transpose. n is dimension of the nonlinear parameter space. The normalization coefficient of the Gaussian function, being irrelevant for the shape of the $p(\mathbf{T})$ landscape, has been dropped. Specifically, when $n=2$, we have

$$\mathbf{C}^{-1}=\begin{pmatrix} a & c \\ c & b \end{pmatrix}, \tag{10}$$

$$p(\mathbf{T}) \propto \exp(-\boldsymbol{\epsilon}^2), \tag{11}$$

with

$$\boldsymbol{\epsilon}^2=\tfrac{1}{2}[a(T_1-\hat{T}_1)^2+2c(T_1-\hat{T}_1)(T_2-\hat{T}_2) + b(T_2-\hat{T}_2)^2], \tag{12}$$

where

$$a=\left.\frac{\partial^2 p}{\partial T_1^2}\right|_{\hat{T}_1,\hat{T}_2}, \quad b=\left.\frac{\partial^2 p}{\partial T_2^2}\right|_{\hat{T}_1,\hat{T}_2}, \quad c=\left.\frac{\partial^2 p}{\partial T_1 \partial T_2}\right|_{\hat{T}_1,\hat{T}_2}. \tag{13}$$

Equation (12) is an ellipse of general form as illustrated in Fig. 2. Comparable to 1D Gaussian distribution in Eq. (5) the covariance matrix $\mathbf{C}$ of a multidimensional Gaussian distribution in Eq. (9) is on the place where the variance $\langle\sigma^2\rangle$ is in Eq. (5). The eigenvalues $\boldsymbol{\Lambda}$ and eigenvectors $\mathbf{Q}$ of the covariance matrix $\mathbf{C}$ determine the length and orientation of principle axes of the ellipse.[14] These two features of the PPDF at the global peak can be used to measure the reliability and inter-relationships of decay times. More specifically, the eigenvalues $\boldsymbol{\Lambda}=\{\lambda_i\}$, served as individual variances, determine the length of principle axes of the characteristic ellipse. By the characteristic ellipse, it means the ellipse with the principle axes $2\sqrt{\lambda_i}$ in length with the orientation determined by the eigenvectors $\mathbf{Q}$. While the cross-correlation coefficient $C_iC_j/\sqrt{C_i^2C_j^2}$ evaluated from the covariance matrix are used to quantify the inter-relationships between decay times.

C. Individual variance estimation

Based on the Gaussian model in Eq. (9) for PPDF around the global peak over the decay time space, covariance matrix in Eq. (9) in general, Eq. (10) when $n=2$ in particular, can be estimated using model-based methods. In this work we draw a set of random samples generated by a normalized Gaussian random process:

$$\int g(\mathbf{T})d\mathbf{T}=1 \tag{14}$$

from the PPDF around the global peak in estimating the expected decay times $\langle\mathbf{T}\rangle$ and the expected covariance matrix $\langle\mathbf{C}\rangle$ using the importance sampling method.[24] The expected decay times can then be approximated by

$$\langle\mathbf{T}\rangle \approx \frac{\Sigma_{r=1}^{R}\mathbf{T}_r q(\mathbf{T}_r)}{\Sigma_{r=1}^{R} q(\mathbf{T}_r)}, \tag{15}$$

with

$$q(\mathbf{T}_r)=\frac{p(\mathbf{T}_r|\mathbf{D},\mathbf{I})}{g(\mathbf{T}_r)}, \tag{16}$$

where R is the total number of samples and $\mathbf{T}_r=\{T_{ir}\}$ with $i=1,2,\ldots$ is the decay time vector drawn from the Gaussian

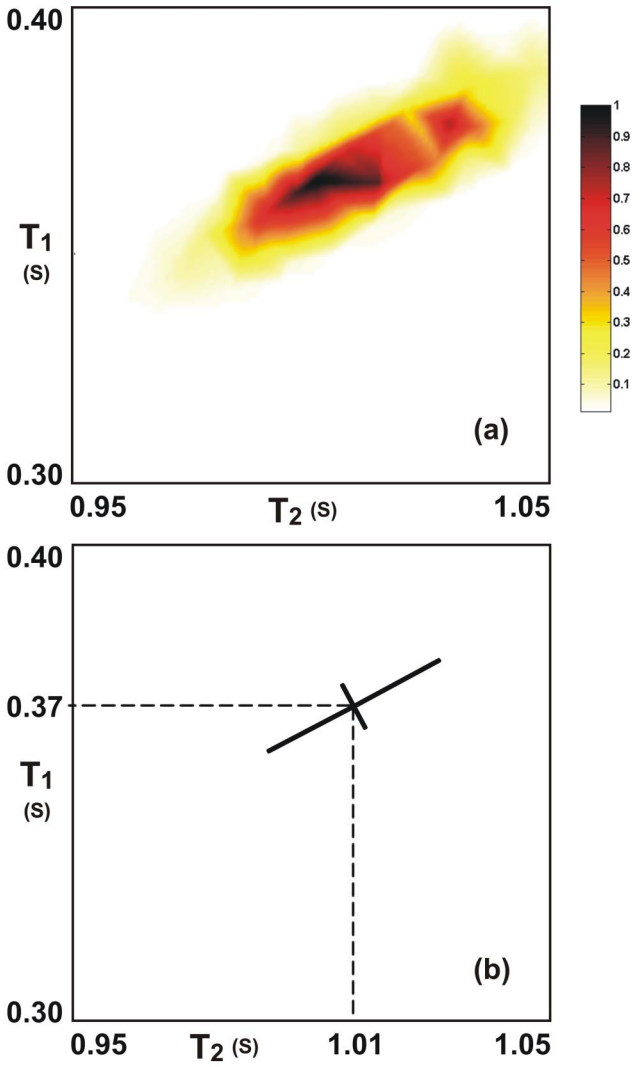


FIG. 3. Estimation of individual variances for the data as shown in Figs. 1(a)–1(c). Important sampling method is applied in estimating the lengths of principle axes of the distribution. (a) Pseudocolor presentation of the density of Gaussian distributed random samples. Figure 4 provides another presentation of the random samples. (b) Two elliptical principle axes estimated for the individual variances along with their final location, associated with the estimates of the expected values of decay times.

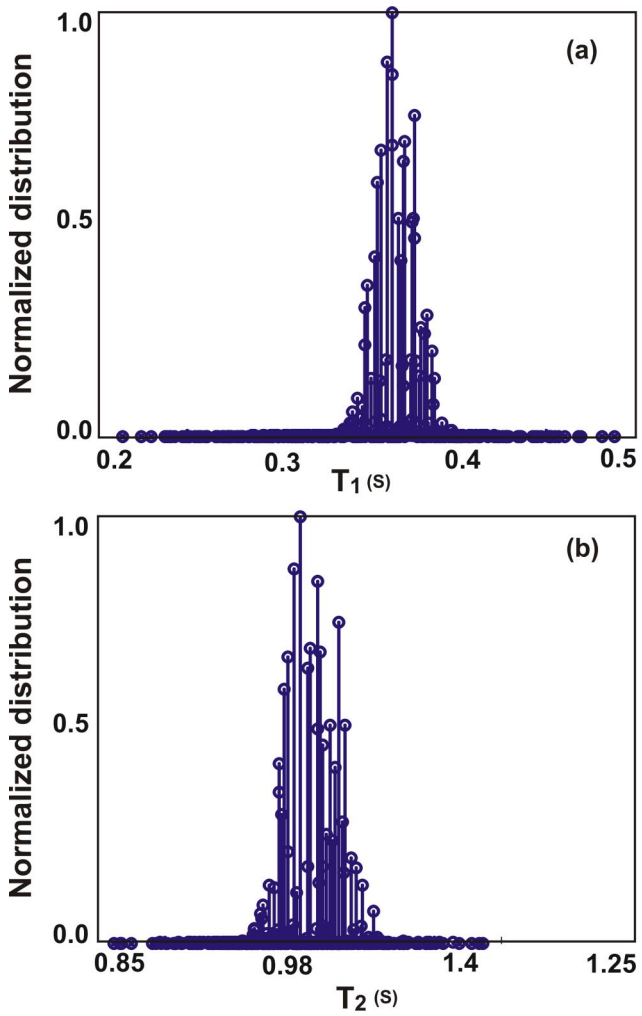


FIG. 4. Peak-normalized marginal distributions $p(\mathbf{T}|\mathbf{D},\mathbf{I})$ of T_1 and T_2 drawn from 512 Gaussian random samples for the data as shown in Figs. 1(a)–1(c). They provide another presentation as shown in Fig. 3(a).

random process at r sampling step. The PPDF $p(\mathbf{T}|\mathbf{D},I)$ is evaluated in form of the student-T distribution given by Eq. (21) of Ref. 4,

$$p(\mathbf{T}|\mathbf{D},I) \propto \left(1 - \frac{m\overline{q^2}}{K\overline{d^2}}\right)^{(m-K)/2}. \tag{17}$$

A random generator associated with the normalized Gaussian process expressed in Eq. (14) generates R random vectors $\mathbf{T}_r$ of n-elements in general, random pairs $\{T_1, T_2\}_r$ for $n=2$, in particular. Using these random decay times in each vector $\mathbf{T}_r$, total R values of $p(\mathbf{T}_r|\mathbf{D},\mathbf{I})$, termed PPDF-samples in this paper, are calculated. Using Eq. (16) each PPDF-sample is scaled by the analytical Gaussian function in Eq. (14) at each decay time vector $\mathbf{T}_r$. Each $q(\mathbf{T}_r)$ value is then sequentially stored in memory along with the associated decay time vector $\mathbf{T}_r$ so that the expected decay time $\langle \mathbf{T} \rangle$ is evaluated straightforwardly using Eq. (15).

The expected covariances can then be approximated by

$$\langle C_{ij} \rangle \approx \frac{\Sigma_{r=1}^{R} (T_{ir} - \langle T_i \rangle)(T_{jr} - \langle T_j \rangle) q(\mathbf{T}_r)}{\Sigma_{r=1}^{R} q(\mathbf{T}_r)}. \tag{18}$$

The stored $q(\mathbf{T}_r)$ values along with $\mathbf{T}_r$ vectors in the memory are reused for estimating the covariance matrix using Eq. (18) with a minor computational load. After estimation of the covariance matrix, its eigenvalues and eigenvectors can be obtained.

Figures 3 and 4 illustrate estimation of the length of principle axes of the PPDF as shown in Figs. 1(a)–(c) from the eigenvalues of the estimated covariance matrix. Their orientation is determined by the estimated eigenvectors. Figure 3(a) illustrates a pseudocolor density representation of 512 samples of $p(\mathbf{T}|\mathbf{D},I)$ drawn by the Gaussian random process for estimating the expected decay times $\langle \mathbf{T} \rangle$ and the expected covariance matrix $\langle \mathbf{C} \rangle$. Figure 3(b) shows the estimated position of two expected decay times, the length, and orientation of the principle axes. Their axis length, estimated from the eigenvalues of the expected covariance matrix, are directly associated with the individual variances $\langle \tau_i^2 \rangle$. In other words, the two crossed lines show the standard deriva-

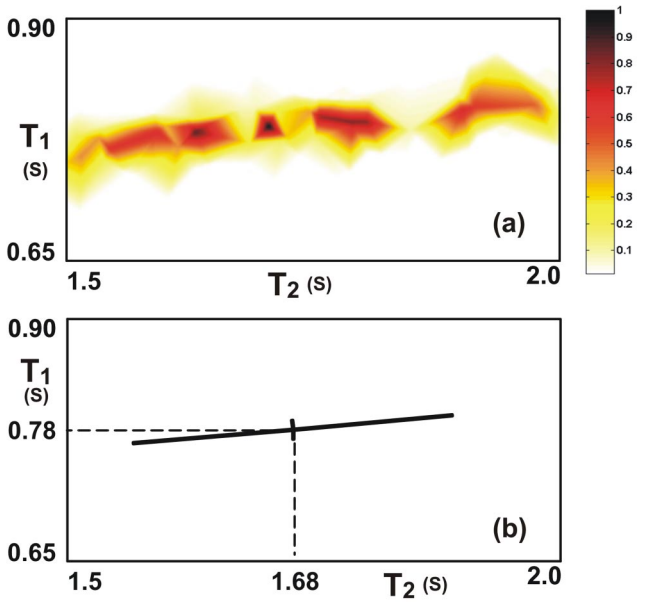


FIG. 5. Estimation of individual variances for the data as shown in Figs. 1(d)–1(f) using the important sampling method. (a) Pseudocolor presentation of the density of Gaussian distributed random samples. Figure 6 provides another presentation of the random samples. (b) Two elliptical principle axes estimated for the individual variances along with their final location, associated with the estimates of the expected values of two decay times.

tions (''error bars'') of the two decay times which are derived directly from the square root of the eigenvalues of the covariance matrix. From the cross point to the line end, one half of the line on each side corresponds to the individual standard derivation $\langle \tau_i \rangle$. The cross point is then associated with two expected decay times. These expected values (decay times and individual standard derivations) are listed in Table II. Figure 4 provides another view of the estimation as shown in Fig. 3(a); it illustrates the marginal distributions of $p(\mathbf{T}|\mathbf{D},I)$ over the decay time space $\{T_1,T_2\}$ using 512 Gaussian random samples in order to estimate the expected values $\langle T_1 \rangle$ and $\langle T_2 \rangle$ and the expected covariance matrix. These are distribution samples of $p(\mathbf{T}|\mathbf{D},I)$ over the decay time space $\{T_1,T_2\}$ using 512 Gaussian random samples, but projected onto the T_1 and T_2 axes, respectively. The marginal distributions as shown in Fig. 4 are normalized with respect to the peak-value (peak-normalization). In a similar fashion, Figs. 5 and 6 illustrate the estimation of the length of the principle axes and their orientation of the PPDF as shown in Figs. 1(d)–1(f). Figure 5(a) illustrates a pseudocolor density representation of 512 samples of $p(\mathbf{T}|\mathbf{D},I)$ drawn by the Gaussian random process. Figure 5(b) shows the estimated position of two expected decay times, the expected standard derivations, and the orientation of the principle axes. The estimates of the expected values (decay times and individual standard derivations) are listed in Table II. Figure 6 illustrates the peak-normalized marginal distributions of $p(\mathbf{T}|\mathbf{D},I)$ over the decay time space $\{T_1,T_2\}$ using 512 Gaussian random samples.

TABLE II. Expected values $\langle T_1 \rangle$, $\langle T_2 \rangle$ and individual (expected) standard derivations $\langle \tau_1 \rangle$, $\langle \tau_2 \rangle$ of the decay times of two different Schroeder's decay functions from experimentally measured impulse responses. ''Union'' stands for the Student Union at the University of Mississippi, ''Church'' for the Church in Pascagoula, Mississippi.

Hall	$\langle T_1 \rangle$ (s)	$\langle \tau_1 \rangle$ (s)	$\langle T_2 \rangle$ (s)	$\langle \tau_2 \rangle$ (s)	Correlation
Church	0.366	0.0103	1.010	0.018	0.854
Union	0.790	0.0157	1.738	0.152	0.80

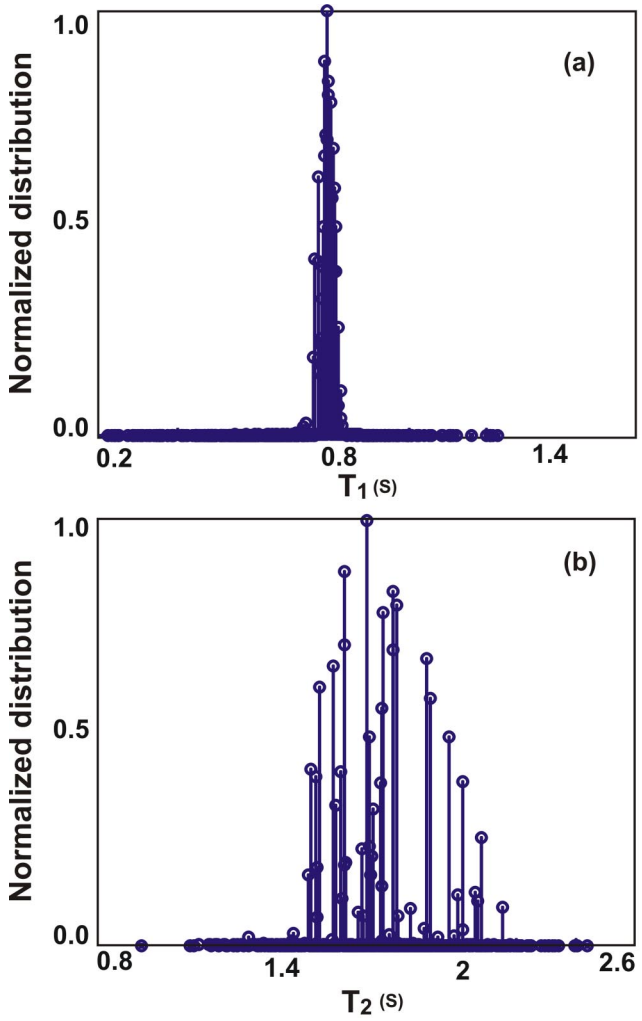


FIG. 6. Peak-normalized marginal distributions $p(\mathbf{T}|\mathbf{D},\mathbf{I})$ of T_1 and T_2 drawn from 512 Gaussian random samples for the data as shown in Figs. 1(a)–1(c). They provide another presentation as shown in Fig. 5(a).

The importance sampling method used in this work is a random procedure. Table III also lists three different runs of the estimation results for the example shown in Figs. 1(a)–1(c) using 256, 512, 1024, and 2048 Gaussian random samples. The estimation results between each single runs may differ from each other. However the estimated final values of the expected decay times and associated individual variances are already in the precision needed in architectural acoustics practice. This example suggests that spending 256/512 or even 2048 samples [PPDF-calculations using Eq. (16)] would be much more efficient than the exhausted mapping of the full PPDFs in that small region using 300 000–640 000 PPDF-calculations as shown in Figs. 1(c) and (f).

For the estimation the given set of Gaussian random samples are drawn within a relative large decay time space as indicated as initial Gaussian assignment in Table III. At the initial stage where a Gaussian distribution needs to be as-

TABLE III. Estimated expected decay times $\langle T_1 \rangle$, $\langle T_2 \rangle$ and their individual standard derivations $\langle \tau_1 \rangle$, $\langle \tau_2 \rangle$, along with the cross-correlation coefficient indicating the dependence of two decay times of the Schroeder's decay function from experimentally measured impulse response in a Church in Pascagula, Mississippi.

	$\langle T_1 \rangle$ (s)	St. derivation $\langle \tau_1 \rangle$ (s)	$\langle T_2 \rangle$ (s)	St. derivation $\langle \tau_2 \rangle$ (s)	Correlation
Initial Gaussian assignment	0.374	0.05	1.02	0.05	0
	0.3661	0.0107	1.0106	0.0181	0.8366
256 samples	0.3676	0.0182	1.0099	0.0119	0.8445
	0.3631	0.0107	1.0109	0.0150	0.8637
	0.3657	0.0094	1.0083	0.0153	0.8426
512 samples	0.3638	0.0108	1.0091	0.0159	0.8394
	0.3660	0.0090	1.0102	0.0167	0.8775
	0.3652	0.0101	1.0101	0.0172	0.8965
1024 samples	0.3659	0.0092	1.0089	0.0167	0.8567
	0.3661	0.0110	1.0103	0.0184	0.8237
	0.3650	0.0102	1.0080	0.0159	0.8902
2048 samples	0.3661	0.0099	1.0095	0.0165	0.8559
	0.3658	0.0100	1.0102	0.0177	0.8936

signed for drawing random samples, the initial values of T_1 and T_2, as estimated using a search algorithm based on Gibbs sampling method in Sec. III A (see Ref. 4), served as initial MAP-values for the Gaussian distribution. An independence between the decay parameters (cross-correlation=0) is initially assumed. Two initial variances for the Gaussian distribution can be roughly estimated using Eq. (13) in terms of a few samples around the initial MAP-values of the decay times. The importance sampling method will yield precise results when the distribution density of the random process in Eq. (14), with which the random samples are generated, are close to the PPDF in Eq. (17).[24] As shown in Table III, however, even a relative rough assignment of the initial MAP-values of decay times and the initial variances to the Gaussian random sample generator will lead to a reasonable precision of the expected values of decay times and their variances using the importance sampling.

The sampling method given in Eqs. (15) and (18) is implicitly a model-based procedure. It assumes an appropriate model of the PPDF around one global peak as given in Eq. (9). Estimation of expected values of decay times and the individual variances using the importance sampling is a moment calculation while the initial MAP values estimated using Gibbs sampling represents a location search algorithm proceeding the importance sampling. So the estimation of expected decay times and their individual variances for reliability analysis in this work is carried out using hybrid sampling methods. When one decay time is dependent on another (correlation>0), any search algorithm will suffer from a long search process.[15,18] The hybrid methods implemented in this work can avoid back-and-forth search across the global extreme over the decay time space when the PPDF landscape is skewed,[15] since a rough assignment for the Gaussian random samples is sufficient as shown by the numerical examples. It implies that an earlier interruption of Gibbs sampling for searching the initial decay times is possible. One further step and/or even iterative steps of the importance sampling, using the updated decay values and individual variances achieved in the previous step, can be conceivably applied to obtain even higher precision of estimation. Since the precision as shown in Table III has already proven sufficient for architectural acoustics practice, the iterative importance sampling is not shown in this work, rather it is highly recommended for other applications.

IV. CONCLUSIONS

Estimation of decay times in acoustically coupled spaces can be accomplished within the Bayesian framework. The reliability of the decay time estimation is primarily determined by the data given a well-established decay model. Both general variance and individual variances are important measures for the reliability of the decay time estimation. They should always be estimated as part of the sound energy decay analysis. They can serve as quality indicators in interpreting the results with quantitative confidence. This paper describes the handy tools for the reliability analysis. Along with the tools for the decay parameter estimation[4] and the decay model selection[5] within Bayesian framework, Bayesian approaches are suitable methods for the decay time evaluation in acoustically coupled spaces.

ACKNOWLEDGMENTS

The authors would like to thank Professor Manfred R. Schroeder for his unremitting interest and inspiration. This paper is dedicated to Professor Manfred R. Schroeder on the occasion of the 40th anniversary of the publication of his integrated impulse response method [J. Acoust. Soc. Am. **37**, 409–412 (1965); aspects of this work have been presented at the 148th ASA Meeting, San Diego, CA (2004). Dr. Dezhang Chu and Dr. Wing T. Chu are also thanked for their valuable discussion and advice. The authors would also like to express their gratitude to Dr. Chris Jaffe and Dr. Rendell R. Torres, who supported this work with enthusiasm, to Donghua Li, who participated in extensive room-acoustic data collection in the scope of his M.S. degree thesis. Furthermore, the authors would like to thank the reviewers for their insightful and constructive comments.

APPENDIX

A substitution of Eq. (6) into Eq. (7) yields

$$\langle \sigma^s \rangle = \frac{\int_0^\infty \sigma^{-K+m+s-1} \exp\left(-\frac{K\overline{d^2} - m\overline{q^2}}{2\sigma^2} \right) d\sigma}{\int_0^\infty \sigma^{-K+m-1} \exp\left(-\frac{K\overline{d^2} - m\overline{q^2}}{2\sigma^2} \right) d\sigma}. \quad \text{(A1)}$$

A change of variable $u=1/\sigma$; $d\sigma=-\sigma^2 du$ and a replacement of the integral limits $\sigma\to 0$; $u\to\infty$ and $\sigma\to\infty$; $u\to 0$ lead the denominator of Eq. (A1) to the form

$$\int_0^\infty u^{2n} \exp(-au^2) du = \frac{1}{2}\Gamma\left(n+\frac{1}{2}\right) a^{-(n+1/2)}, \quad \text{(A2)}$$

with $a=(K\overline{d^2}-m\overline{q^2})/2$ and $n=(K-m-s-1)/2$ the denominator becomes

$$\int_0^\infty \sigma^{-K+m+s-1}\exp\left(-\frac{K\overline{d^2}-m\overline{q^2}}{2\sigma^2}\right)d\sigma = \frac{1}{2}\Gamma\left(\frac{K-m-s}{2}\right)a^{-(K-m-s)/2}. \quad \text{(A3)}$$

The numerator can be determined in similar fashion so that

$$\langle\sigma^s\rangle=\Gamma\left(\frac{K-m-s}{2}\right)\Gamma\left(\frac{K-m}{2}\right)^{-1}\left(\sqrt{\frac{K\overline{d^2}-m\overline{q^2}}{2}}\right)^s. \quad \text{(A4)}$$

[1] M. Ohta, H. Yamada, and H. Iwashige, ''A system-theoretical evaluation method for the reverberation time of an acoustically couled room system,'' J. Acoust. Soc. Jpn. (E) **16**, 137–145 (1995).

[2] J. S. Anderson, M. Bratos-Anderson, and P. Doany, ''The acoustics of a large space with a repetitive pattern of coupled rooms,'' J. Sound Vib. **208**, 313–329 (1997).

[3] V. Valeau, J. Picaut, A. Sacaut, and A. Billon, ''Simulation of the acoustics in coupled spaces by numerical resolution of a diffusion equation,'' *Proc. ICA'04*, Vol. I, pp. 483–498.

[4] N. Xiang and P. M. Goggans, ''Evaluation of decay times in coupled spaces: Bayesian parameter estimation,'' J. Acoust. Soc. Am. **110**, 1415–1424 (2001).

[5] N. Xiang and P. M. Goggans, ''Evaluation of decay times in coupled spaces: Bayesian decay model selection,'' J. Acoust. Soc. Am. **113**, 2685–2697 (2003).

[6] J. E. Summers, R. R. Torres, and Y. Shimizu, ''Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics,'' J. Acoust. Soc. Am. **116**, 958–969 (2004).

[7] D. T. Bradley and L. M. Wang, ''Relating double slope decay in coupled volumes to volume ratio, absorption ratio, and aperture size,'' J. Acoust. Soc. Am. (to be published).

[8] M. R. Schroeder, ''New method of measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412 (1965).

[9] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).

[10] F. Jacobsen, ''A note on acoustic decay measurements,'' J. Sound Vib. **115**, 165–170 (1987).

[11] M. Vorländer and H. Bietz, ''Comparison of methods for measuring reverberation time,'' Acustica **80**, 205–215 (1994).

[12] A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, ''Uncertainties of measurements in room acoustics,'' Acustica **81**, 344–355 (1995).

[13] D. R. Morgan, ''A parametric error analysis of the backward integration method for reverberation time estimation,'' J. Acoust. Soc. Am. **101**, 2686–2693 (1997).

[14] D. S. Sivia, *Data Analysis: A Bayesian Tutorial* (Clarendon, Oxford, 1996).

[15] M. D. Collins and L. Fishman, ''Efficient navigation of parameter landscapes,'' J. Acoust. Soc. Am. **98**, 1637–1644 (1995).

[16] S. Dosso, ''Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach,'' J. Acoust. Soc. Am. **111**, 129–142 (2002).

[17] S. Dosso and P. Nielson, ''Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data,'' J. Acoust. Soc. Am. **111**, 143–159 (2002).

[18] T. B. Neilson, ''An iterative implementation of rotated coordinates for inverse problems,'' J. Acoust. Soc. Am. **113**, 2574–2586 (2003).

[19] N. Xiang, ''Evaluation of reverberation times using a nonlinear regression approach,'' J. Acoust. Soc. Am. **98**, 2112–2121 (1995).

[20] L. Faiget, C. Legros, and R. Ruiz, ''Optimization of the impulse response length: Application to noisy and highly reverberant rooms,'' J. Audio Eng. Soc. **46**, 741–750 (1998).

[21] H. Jeffreys, *Theory of Probability*, 3rd ed. (Oxford University Press, London, 1961).

[22] G. L. Bretthorst, ''Bayesian Analysis. I. Parameter Estimation Using Quadrature NMR Models,'' J. Magn. Reson. (1969-1992) **88**, 533–551 (1990).

[23] G. L. Bretthorst, ''Bayesian Analysis. IV. Noise and Computing Time Consideration,'' J. Magn. Reson. (1969-1992) **93**, 369–394 (1991).

[24] J. J. K. O'Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer, Berlin, 1996).

Prediction of reverberation time and speech transmission index in long enclosures

Kai Ming Li[a] and Pou Man Lam
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

(Received 17 January 2005; revised 18 March 2005; accepted 19 March 2005)

It is known that the sound field in a long space is not diffuse, and that the classic theory of room acoustics is not applicable. A theoretical model is developed for the prediction of reverberation time and speech transmission index in rectangular long enclosures, such as corridors and train stations, where the acoustic quality is important for speech. The model is based on an image-source method, and both acoustically hard and impedance boundaries are investigated. An approximate analytical solution is used to predict the frequency response of the sound field. The reverberation time is determined from the decay curve which is computed by a reverse-time integration of the squared impulse response. The angle-dependence of reflection coefficients of the boundaries and the change of phase upon reflection are incorporated in this model. Due to the relatively long distance of sound propagation, the effect of atmospheric absorption is also considered. Measurements of reverberation time and speech transmission index taken from a real tunnel, a corridor, and a model tunnel are presented. The theoretical predictions are found to agree well with the experimental data. An application of the proposed model has been suggested.
[DOI: 10.1121/1.1904443]

PACS numbers: 43.55.Ka, 43.55.Br, 43.55.Gx, 43.55.Hy [NX] Pages: 3716–3726

I. INTRODUCTION

Among various acoustic indices used in room acoustics, reverberation time plays an important role in the design of a room for speech or music, while the speech transmission index has been widely accepted as an indicator of speech intelligibility. These two parameters are crucial, for example, in the design of public address systems in long enclosures, such as underground stations and building corridors. Increasing attention has been paid to the investigation of the physical phenomena of sound propagation in long enclosures, where the assumption of a diffuse sound field in the classic theory of room acoustics[1] is not strictly applicable.

Numerical formulations based on energy approach have been developed to predict the reverberation time[2–4] and sound transmission index[5] in long enclosures. The phase information of the reflected waves was generally ignored, and the reflection coefficients were averaged over the frequency band of interest. Significant discrepancies between the predicted and measured data were reported, especially at low frequency, because the interference effects were not included in these earlier models. The surface absorption was assumed to be angle-independent, which is unlikely to be the case in practice for sound absorption materials. An alternative way to evaluate the reverberation time and speech transmission index was to deduce them from the squared impulse response.[6] The sound field in a long enclosure is first computed in the frequency domain, followed by the application of an inverse Fourier transform to obtain the impulse response. Suh and Nelson[7] showed that when the phase information of the waves was included, the model gave a more accurate prediction. António *et al.*[6] and Dance *et al.*[8] also pointed out the importance of interference effects in an enclosed space. However, in the numerical model by António *et al.*,[6] the absorption coefficients used were still band-averaged values and independent of incidence angle. In their work, only the cases of 0° and 180° phase shift upon reflection were studied.

Recently, Li and Iu presented a ''coherent'' model, or the complex image source model, for the prediction of sound propagation in street canyons,[9] long enclosures[10] and tunnels.[11] The model was based on an analytic Green's function, which allowed a variation of the reflection factors with the angle of incidence. The total sound pressure was obtained by summing the contribution from each image. The mutual interference effects of the direct and reflected sound waves were considered. It was shown that the coherent model[12] gave a more accurate prediction than the incoherent model. Unlike the coherent model, the plane wave reflection coefficient was used instead of the spherical wave reflection coefficient in Suh and Nelson's model.[7] They pointed out that their model gave a good approximation, except for the prediction of sound fields of near grazing incidence. When the propagation distance increases, it becomes more and more likely for the propagation of near-grazing waves in long enclosures. Consequently, the spherical wave reflection coefficient should be used instead of the plane wave reflection coefficient.

In this paper, the coherent model is extended to predict the reverberation time (RT30 and EDT) and speech transmission index (STI) in a long enclosure with a rectangular cross section. The difference between the use of spherical and plane wave reflection coefficients in the computation will be studied in a numerical simulation. They are presented in Sec. II. Measurements have been carried out in a road traffic tunnel, a corridor in a building and a model corridor to explore

[a]Electronic mail: mmkmli@polyu.edu.hk

 0001-4966/2005/117(6)/3716/11/$22.50

the validity of the proposed theoretical formulation. The experimental results and their comparisons with the theoretical models are given in Sec. III. Numerical predictions of the Speech Transmission Index in a long enclosure were shown in Sec. IV to demonstrate the usefulness of the proposed model. Finally, some concluding remarks are offered in Sec. V.

II. THEORY

A rectangular long enclosure is modelled as two parallel vertical walls and two parallel horizontal planes. Based on the image source method, the boundaries are assumed to be geometrically reflective, so that the effect of diffusion is ignored. When a point source is located in a rectangular enclosure, an infinite number of virtual image sources are formed. These sources can be identified geometrically.

A. Ray model: Incoherent summation

According to Kang's model,[2] the steady-state sound pressure level is given by

$$\mathrm{SPL}_d = 10 \log \sum_{\Delta t} 10^{\mathrm{SPL}_d(t)/10} - \mathrm{SPL}_{\mathrm{ref}}, \tag{1}$$

where

$$\mathrm{SPL}_d(t) = 10 \log \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{K e^{-a d_n} [(1-\alpha)^n]}{d_n^2}, \tag{2}$$

d_n is the path length from the image source n to the receiver, K is a constant relating to the sound power of the source, α is the absorption coefficient of the boundaries, a is the air absorption coefficient, and $\mathrm{SPL}_{\mathrm{ref}}$ is the reference sound pressure level. If the four boundaries have different levels of absorptiveness, the term in the squared brackets will be replaced with a combined reflection factor determined by the four absorption coefficients and the reflection order. A decay curve can be generated by a reverse-time integration of Eq. (2), and the reverberation time is then deduced from the curve. It can be seen from Eq. (2) that the mutual interference effects of the direct and reflected waves are ignored. This approach is referred to as the ''incoherent'' model here. The absorption coefficients of a boundary surface for different frequency bands used in this model are averaged band values. They do not vary with the angle of incidence, thus cannot give a good representation of the true physical absorptiveness of the surface. The phase change upon reflection from a boundary is also ignored.

B. The coherent model

In the coherent model,[11] the total sound field due to monopole of unit strength can be computed by summing all contributions coherently to yield

$$P(\omega) = \frac{1}{4\pi} \sum_{n=0}^{\infty} Q_n \frac{e^{ikd_n}}{d_n}, \tag{3}$$

where $k(\equiv \omega/c)$ is the wave number, ω is the angular frequency and c is the speed of sound in air, d_n is the distance of the image source n from the receiver, and Q_n is the combined complex wave reflection coefficient associated with the image source n. At each interaction with a boundary surface, the spherical wave reflection coefficient Q is determined according to[13]

$$Q = R_p + (1 - R_p) F(w), \tag{4}$$

where R_p and $F(w)$ are the plane wave reflection coefficient and the boundary loss factor respectively. The computation of the terms is given in Ref. 11 and not repeated here. As mentioned in the Introduction, the spherical wave reflection coefficient was replaced by the plane wave reflection coefficient in the numerical model by Suh and Nelson.[6] In other words, the second term in Eq. (4) was ignored in their model. This term accounts for the ''ground wave'' of the reflected sound. The effect of this term on the computation will be discussed in Sec. II D. The time-dependent factor, $e^{-i\omega t}$, is understood and suppressed throughout.

The impulse response of sound pressure is computed by taking the inverse Fourier transform of the frequency response

$$p(t) = \mathrm{IFFT}(P(\omega)). \tag{5}$$

The time interval and the length of the FFT depend on the frequency of interest. For octave band analysis, the frequency response is first filtered before applying the inverse Fourier transform. Alternatively, if the source generates a known band limited time signal instead of white noise, the frequency response is multiplied by the Fourier transform of the source strength signal. The impulse response is then obtained by taking the inverse Fourier transform of the product.

A curve showing the decay of sound energy after the cessation of the source can be generated by a reverse-time integration of the squared impulse response, known as the Schroeder approach,[14]

$$E(t) = 10 \log \left\{ \int_t^{\infty} [p(\tau)]^2 d\tau \Big/ \int_0^{\infty} [p(\tau)]^2 d\tau \right\}. \tag{6}$$

RT30 is determined using the rate of decay given by the linear regression of the decay curve from a level of 5 dB below the initial level of 35 dB below, while the early decay time (EDT) is obtained from the initial 10 dB of the decay. Other indices related to speech intelligibility, such as clarity and definition, can also be derived from the impulse response.

The number of reflections used in the computation is estimated by considering the required duration (T_D) for the impulse response in the time domain. The duration is determined by the inverse of the frequency increment ($T_D = 1/\Delta f$) used in the numerical analysis. These two parameters are chosen according to the configuration of the experimental set-up.

The critical frequency where the coherent model should be used instead of the incoherent model was introduced by Li and Iu.[12] Assuming the source-receiver distance $L \gg 0$, the critical frequency is determined by

$$f_c \sim 2c(L^2 + A)/AL, \tag{7}$$

where c is the speed of sound in air, and A is the cross-sectional area of the long enclosure.

C. Speech transmission index (STI)

The speech transmission index is widely accepted as an indicator of speech intelligibility. It is based on the modulation transfer function as described by Houtgast and Steeneken.[15] When both the effect of reverberation and ambient background noise are taken into account, the modulation transfer function $m(F)$ is given in a general form as

$$m(F)=\frac{\int_0^\infty e^{2\pi jF\tau}[p(\tau)]^2 d\tau}{\int_0^\infty [p(\tau)]^2 d\tau}[1+10^{(-\mathrm{S/N})/10}]^{-1}, \tag{8}$$

where S/N is the signal-to-noise ratio in dB at the receiver's position and F is the modulation frequency. It can be seen from Eq. (8) that the contribution of ambient background noise to the transfer function is inversely proportional to the signal-to-noise ratio. When the signal-to-noise ratio is sufficiently high, the contribution of the term in squared brackets of Eq. (8) is negligible, and the formulation becomes

$$m(F)=\frac{\int_0^\infty e^{2\pi jF\tau}[p(\tau)]^2 d\tau}{\int_0^\infty [p(\tau)]^2 d\tau}. \tag{9}$$

The modulation transfer function is computed with 14 modulation frequencies, from 0.63 Hz to 12.5 Hz, in one-third octave intervals. These are then converted into a single index, the speech transmission index (STI), which is defined by[16]

$$\mathrm{STI}=\sum_{f=125\ \mathrm{Hz}}^{8\ \mathrm{kHz}} w_f \mathrm{STI}_f, \tag{10}$$

where w_f is the frequency weighting factor. From 125 Hz to 8 kHz (octave), w_f is equal to 0.13, 0.14, 0.11, 0.12, 0.19, 0.17, and 0.14, respectively. The parameter STI_f is the transmission index at each of the seven octave bands. The computation of STI_f is described, for example, in Ref. 17. In the experimental validation of the proposed model in Sec. III, the measured and predicted STI_f at different octave bands, together with the overall STI, will be compared.

D. Effect of the approximation of the spherical wave reflection coefficient with the plane wave coefficient

The effect of the replacement of the spherical wave reflection coefficient with the plane wave coefficient on the numerical prediction of EDT, RT30, and STI is studied in this section. It is analyzed with an imaginary long rectangular enclosure of width 6 m and height 4 m. The enclosure is assumed to be infinitely long, meaning there is no reflection from the walls at either end. The vertical walls of the enclosure are perfectly hard. The horizontal planes are impedance boundaries. They share the same level of absorptiveness as characterized by the one-parameter model[18] based on the effective flow resistivity, σ_e, of the boundary surfaces. Four values of σ_e are used in the computation: 100 kPa s m^{-2}, 500 kPa s m^{-2}, 1000 kPa s m^{-2}, and 10 000 kPa s m^{-2}. They are used to simulate the different levels of absorptiveness of the horizontal boundaries.

The different locations of the point source and receiver are listed in Table I. In case 1, both the source and receiver are located at the center of the cross section, with source–receiver distances of 5 m, 20 m, and 50 m, respectively. The source is then moved to a higher position, which is 3.8 m from the lower boundary in case 2. In the third case, the receiver is also moved to a height of 3.8 m, at the same level as the source. The source and receiver are separated by 5 m, 20 m, and 50 m in each case. Computations of EDT, RT30, and STI are made at each source–receiver combination, with the four different values of effective flow resistivity as mentioned before. In order to illustrate the difference in computation of the two reflection coefficients, data are presented as relative error, Δ, which is given by

TABLE I. Location of the source and receiver in the rectangular enclosure.

Case	Source height	Receiver height	Source–receiver distance
1a	2 m	2 m	5 m
1b	2 m	2 m	20 m
1c	2 m	2 m	50 m
2a	3.8 m	2 m	5 m
2b	3.8 m	2 m	20 m
2c	3.8 m	2 m	50 m
3a	3.8 m	3.8 m	5 m
3b	3.8 m	3.8 m	20 m
3c	3.8 m	3.8 m	50 m

$$\Delta_{\mathrm{RT30}}=\frac{|\mathrm{RT30}(Q)-\mathrm{RT30}(R_p)|}{\mathrm{RT30}(Q)}\times 100\%, \tag{11a}$$

$$\Delta_{\mathrm{EDT}}=\frac{|\mathrm{EDT}(Q)-\mathrm{EDT}(R_p)|}{\mathrm{EDT}(Q)}\times 100\%, \tag{11b}$$

$$\Delta_{\mathrm{STI}}=\frac{|\mathrm{STI}(Q)-\mathrm{STI}(R_p)|}{\mathrm{STI}(Q)}\times 100\%, \tag{11c}$$

where $\mathrm{EDT}(Q)$, $\mathrm{RT30}(Q)$, and $\mathrm{STI}(Q)$ are the corresponding parameters computed with the spherical wave reflection coefficient, Q, and $\mathrm{EDT}(R_p)$, $\mathrm{RT30}(R_p)$ and $\mathrm{STI}(R_p)$ are those computed with the plane wave reflection coefficient, R_p. Comparisons are made in relative "error" so that the effect of the use of Q and R_p on the predicted parametric values of EDT, RT30, and STI can be assessed.

Figures 1, 2, and 3 show the relative error of EDT, RT30, and STI, respectively. The results were computed at different frequency bands, source–receiver positions and absorptiveness of the boundaries. These three factors are believed to be associated with the effect of the spherical wave reflection coefficient [cf. Eq. (4)] on the prediction of reverberation time and speech transmission index. The results of Δ_{EDT} and Δ_{RT30} are presented in one-third octave bands, and those of Δ_{STI} are presented in octave bands. In Fig. 1, for the Δ_{EDT} bars that extend to the upper boundary line, the actual values are over 50% and marked beside the bars.

It is shown clearly that the relative error decreases with the increase of frequency. As mentioned before, four different values of effective flow resistivity are used in the calculation to simulate the different impedance of the horizontal boundaries. The relative error is higher when the boundaries are more absorptive in most cases. It is most significant when

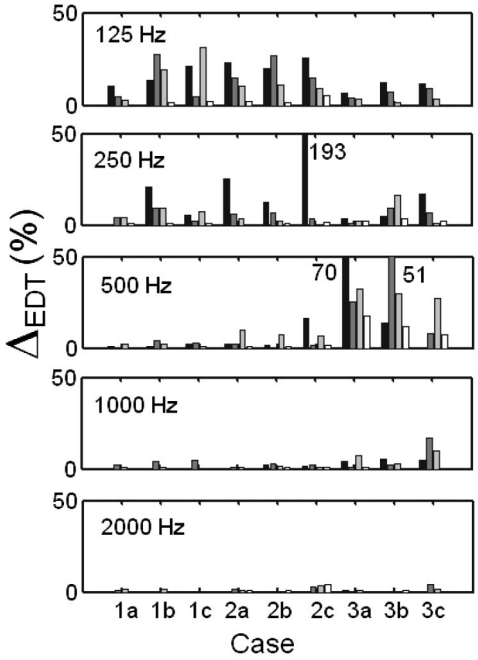


FIG. 1. Relative error in the use of plane wave reflection coefficients to predict EDT at different source–receiver geometry. Results were computed with effective flow resistivities of 100 kPa s m^{-2} (black), 500 kPa s m^{-2} (dark gray), 1000 kPa s m^{-2} (light gray), and 10 000 kPa s m^{-2} (white).

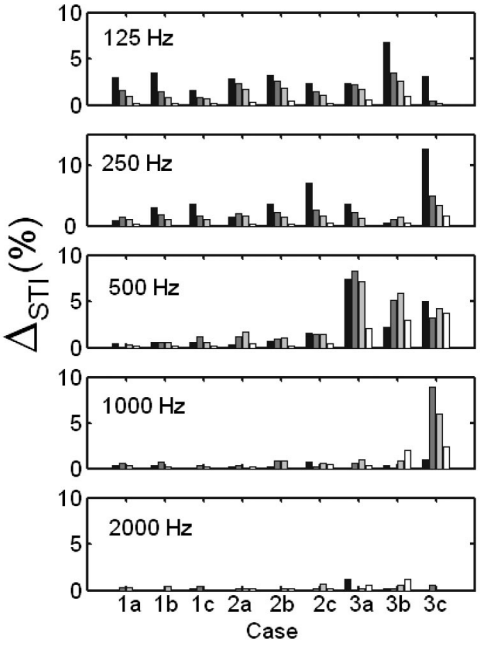


FIG. 3. Relative error in the use of plane wave reflection coefficients to predict STI at different source–receiver geometry. Results were computed with effective flow resistivities of 100 kPa s m^{-2} (black), 500 kPa s m^{-2} (dark gray), 1000 kPa s m^{-2} (light gray), and 10 000 kPa s m^{-2} (white).

the effective flow resistivity changes from 100 kPa s m^{-2} to 500 kPa s m^{-2}. There are exceptions, for example, in the prediction of EDT at 500 Hz in case 3. The initial stage of the decay process may be affected by a combined effect of the interference phenomenon and the locations of the source and receiver.

We remark that the existence of ground wave term has long been identified in the propagation of sound above a ground surface at long ranges.[19] Each of the 4 boundary surfaces of the rectangular enclosure has its own ''ground'' wave term. Due to the existence and significance of these ground wave components, the relative error is expected to increase with the source–receiver distance, and especially when the source and receiver are positioned near a boundary. This can be observed, for example, in the prediction of Δ_{EDT} and Δ_{RT30} at 125 Hz in case 1, Δ_{RT30} at 250 Hz in cases 1 and 2, and that of Δ_{STI} at 500 Hz in case 3. However, the trend is not consistent in other cases. This may be caused by a combined effect of the source–receiver distance and the reflection order of contributing images. When the source–receiver distance increases, the contribution of the direct wave becomes smaller. The sound field is then dominated by the reverberant field. The incident angle of an image increases with its reflection order, and the sound propagation is not near-grazing anymore.

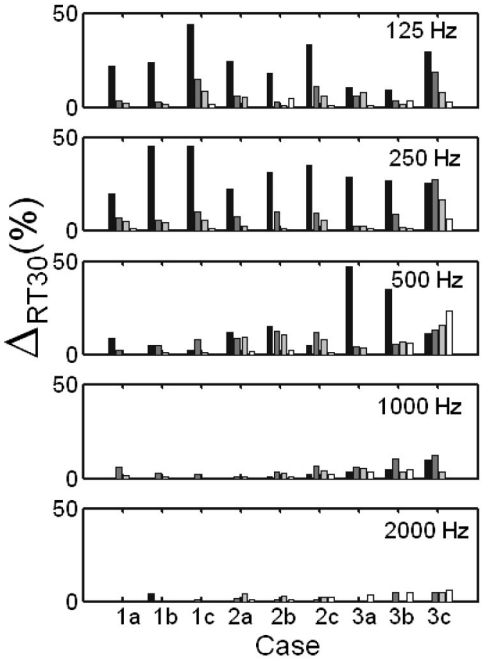


FIG. 2. Relative error in the use of plane wave reflection coefficients to predict RT30 at different source–receiver geometry. Results were computed with effective flow resistivities of 100 kPa s m^{-2} (black), 500 kPa s m^{-2} (dark gray), 1000 kPa s m^{-2} (light gray), and 10 000 kPa s m^{-2} (white).

It is noticed that the effect on the prediction of reverberation time is bigger than that of speech transmission index. Although there does not seem to be an exact rule to identify the trend of the overall effect, the data in Figs. 1–3 show that the use of plane wave reflection coefficients to approximate the spherical one does alter the prediction of the reverberation time and speech transmission index. The difference can be over 100% at certain frequency and boundary conditions. The spherical wave reflection coefficient is, therefore, preferred in the prediction model, especially in the case of sound propagation in a long enclosure.

III. EXPERIMENTAL VALIDATION

A. Tunnel: Long enclosure with perfectly hard boundaries

The Western Harbor Tunnel is a 2 km three-lane road tunnel in Hong Kong's Victoria Harbor. It has a rectangular cross section of nominal width of 12.5 m and height of 5.8 m. The walls and ground are made of concrete, with relatively flat and smooth surfaces. Raised walkways adjacent to the vertical walls and other scattering surfaces hung from the ceiling are found.

A Tannoy Superdual B475 loudspeaker and a Tannoy T300 loudspeaker were used together as a single sound source generating white noise, with the T300 placed on top of the B475. There were other maintenance activities in the tunnel throughout the period of our measurements. This led to a typical background noise level of about 70 dB(A). In our preliminary measurement, it was found that the measured sound pressure level due to the loud speakers reached a minimum level of about 105 dB(A) within a horizontal separation of about 150 m in the tunnel. Hence, the maximum source–receiver distance was limited to 150 m in the present set of measurements. Limiting the maximum source–receiver separation, we guaranteed that the lowest equilibrium sound level at measuring positions was at least 15 dB above the background noise level which was required for the measurement of STI and EDT. A PC-based maximum length sequence system analyzer (MLSSA) (Ref. 20) was used both as the signal generator for the source and as the analyzer for subsequent data processing. Its postprocessing functions calculate most of the acoustical parameters, including the reverberation time and speech transmission index, from the measured impulse response. The same technique was used by Kang[2] and Iu[10] to carry out site measurements in Hong Kong. The receiver used in the measurement was a B&K prepolarized diffuse field $\frac{1}{2}$ in. condenser microphone of type 4942, fitted with a B&K Deltatron preamplifier of type 2671. Data were collected in one-third octave bands.

The source was located at the center of the width of the cross section, i.e., 6.25 m from both vertical walls. The receiver was located at different positions along the centerline (i.e., 6.25 m from the walls) and the offset line (i.e., 9.85 m from one of the walls). The height of the receiver was 1.2 m. Hereafter, the source–receiver distance refers to the horizontal distance along the *y*-axis unless otherwise stated. It ranged from 5 m to 150 m. The average background noise during the measurement was about 70 dB(A). In order to ensure that there was sufficient room for the sound level to decay, and that the results were not affected by the background noise, only the early decay time was measured. With the same level of sound power used in the measurement of STI, the signal-to-noise ratio was sufficiently high in each octave band, and its contribution to the STI was ignored in the computation.

Separate measurements were conducted to characterize the impedance of the wall surfaces and the ground. It was found that they could essentially be treated as perfectly hard surfaces in the numerical models. This is consistent with previous studies by Li and Iu in the same tunnel. The ceiling is assumed to be perfectly hard as well. For the incoherent model, the absorption coefficients (α) were set to 0 in order to simulate the ideal case of perfectly reflecting boundaries. A clear comparison between the coherent and incoherent models can then be made. Setting $\alpha=0$ in Eq. (2), we can generate a set of the decay curves from which the corresponding EDT and STI can be obtained. The effect of atmospheric absorption has been included in the computation. The absorption coefficients for atmospheric absorption are determined according to the formulation given in Ref. 21.

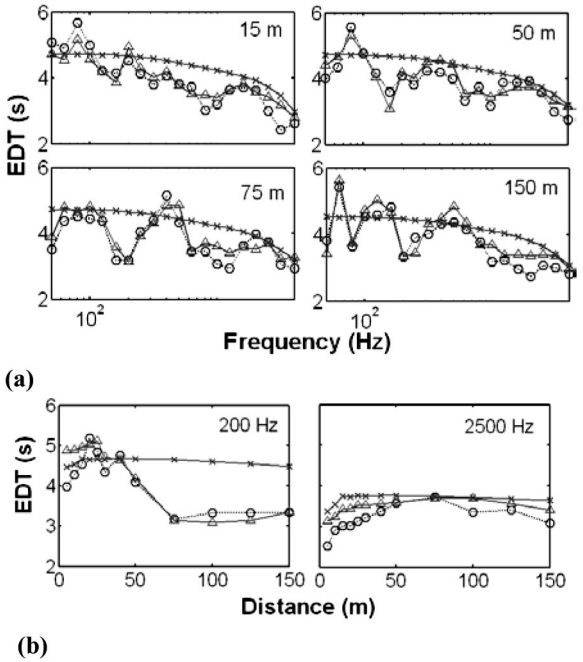


FIG. 4. Comparison of measured and predicted EDT in the tunnel. Circles: measurement; triangles: coherent prediction; crosses: incoherent prediction.

For the coherent prediction, the sound field is first computed in the frequency domain with Eq. (3). Since the boundaries are assumed to be perfectly hard, the term Q_n in Eq. (3) is equal to 1. The results are then digitally filtered by Butterworth IIR bandpass filter. The same type of filter is used in the MLSSA system. The impulse response in the corresponding one-third octave band is obtained by a Fourier transform of the filtered results. A decay curve can be generated by a reverse-time integration of the squared impulse response, as in Eq. (6). The EDT is obtained from the first 10 dB of the decay of sound pressure. The STI is determined from the impulse response as given in the previous section.

Both the EDT and STI were measured at different distances along the centerline and the offset line. It was found that the predictions with the coherent model agreed well with the experimental data, whether the receiver was located on the centerline or the offset line. For brevity of illustration, only the comparison of measured and predicted EDT on the centerline, and that of the STI on the offset line, are presented. Figure 4(a) shows the comparison of EDT spectra at source–receiver separations of 15 m, 50 m, 75 m, and 150 m on the centerline. It can be seen that the incoherent model gives the general trend of the EDT only, while the coherent model provides a better prediction. For example, when the

source and receiver are separated by 75 m, there are dips at 200 Hz and 630 Hz, and a peak at 400 Hz. The coherent model is able to predict the fluctuations at different frequency bands. The incoherent model overestimates by more than 1.5 s at 200 Hz, and 1 s at 630 Hz. At higher frequencies, e.g. over 2000 Hz, the difference between predictions by the two models become smaller.

Figure 4(b) shows how the EDT varies along the length of the tunnel at 2 typical frequencies of 200 and 2500 Hz. At 200 Hz, the EDT fluctuates along the length of the tunnel due to the interference effect of the contributory sound rays. On the other hand, the incoherent model is unable to predict the reduction in EDT at a distance over 50 m at 200 Hz. In fact, it overestimates the EDT for over 1 s at distances beyond 75 m from the source. The coherent and incoherent models give similar predictions in the one-third octave band centered at 2500 Hz, within a difference of 0.5 s. This is in accordance with the findings by António *et al.*,[8] who suggest that the interference effects are significant in low frequency bands. The coherent predictions generally match the measurements better than the incoherent predictions, especially in the low frequency range. For the prediction of the EDT, the average error of the coherent model was 0.3 s, and that of the incoherent model was 0.6 s. We note that the prediction results might match the measurements better at some positions by changing the parametric values of the absorption coefficients in the incoherent model. However, it still cannot predict the fluctuations due to interference effects as predicted by the coherent model as shown in Fig. 4(b).

In the next set of measurements, the receiver was located along the offset line. The observations from comparisons of measured and predicted EDT are similar to the case when the receiver was located at the centerline. In order to show that the coherent model is also valid in this receiver location, the comparisons of measured and predicted STI are shown in Fig. 5. The interference effects are observed in the lower frequency bands, where the coherent model gives a better prediction than the incoherent model. For the overall value of STI, both models give similar predictions, because the overall STI, defined in Eq. (10), is a weighted average of the values at seven octave bands. The incoherent model overestimates the STI at 8000 Hz, but underestimates it at other frequency bands, as shown in Fig. 5. The average errors in predicting the STI were 0.02 and 0.04 for the coherent and incoherent models, respectively. By a numerical weighted average of the values at different octave bands, the incoherent model seems to give a reasonably good prediction of the overall STI, as the corresponding discrepancies are cancelled out mathematically. In this way, the overall STI values can only give a general idea of the speech intelligibility in an enclosure. In order to have a better understanding of the effect of the enclosure on the performance of speech, the STI at different octave bands, of which the coherent model can give a satisfactory prediction, should be investigated as well. Corresponding acoustic treatments can then be applied to improve the speech intelligibility.

It can be seen from Figs. 4 and 5 that the predictions by the coherent model agree reasonably well with the field measurements in most cases. However, some discrepancies between the experimental data and theoretical predictions have been found. This is probably due to the scattering of sound from the raised walkways adjacent to the vertical walls and other scattering objects hung from the ceiling. The fact that the tunnel cross section is not a true rectangular shape also affects the results. Other factors include the use of loud speakers as a point source, and the assumption of a perfectly hard ceiling. It is important to show that the model also works in enclosures with impedance boundaries, which will be discussed in the next section.

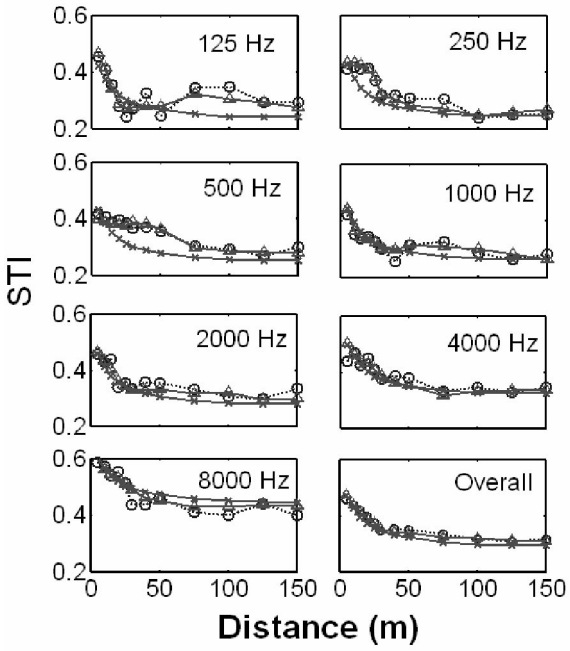


FIG. 5. Comparison of measured and predicted STI, receiver located along the offset line. Circles: measurement; triangles: coherent prediction; crosses: incoherent prediction.

B. Corridor: Long enclosure with two impedance boundaries

Measurements of RT30 and STI were conducted in a corridor in the Department of Mechanical Engineering, the Hong Kong Polytechnic University. The length, width, and height of the corridor are 35.6 m, 1.53 m, and 2.45 m, respectively. The Tannoy T300 loudspeaker was again used as the sound source, generating white noise. The same set of equipment mentioned in the previous section was used, except for the loud speaker. The ground of the corridor was covered with a carpet. The ceiling was made up of perforated metallic panels, embedded with fiberglass.

Separate measurements of excess attenuation (EA) were again made for the characterization of the impedance of the corridor surfaces. A Tannoy driver fitted with a tube of length 1 m was used as a point source. The excess attenuation is defined as the ratio of the total sound field of sound propagation along an impedance surface, P_T, to the direct sound field, P_d,

$$\mathrm{EA} = 20\log(P_T/P_d). \tag{12}$$

The source and receiver were located 0.5 m from each other and at a distance of 0.2 m from the surface of interest. The

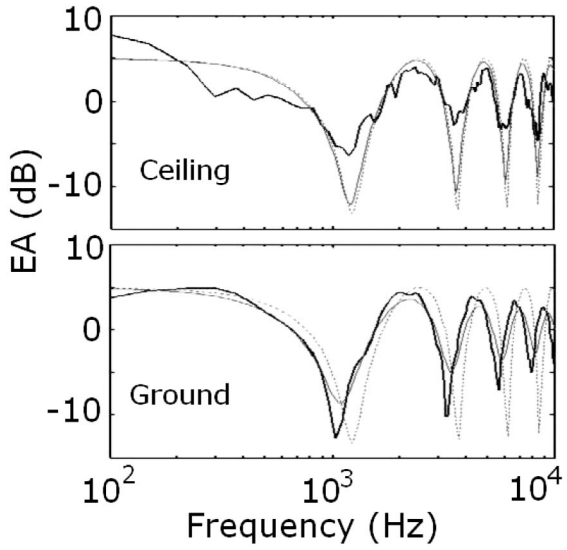


FIG. 6. Plot of excess attenuation against frequency for the characterization of surfaces in the corridor. The thick solid line is the measurement, thin solid line is the prediction with one-parameter model, and dotted line is the prediction with the assumption of a perfectly hard surface.

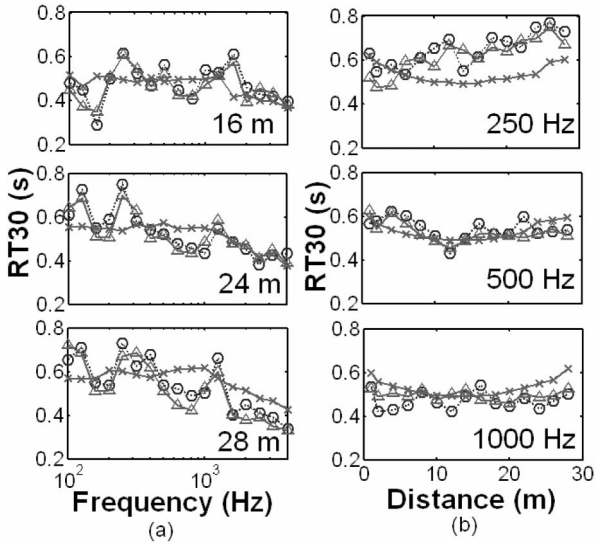


FIG. 7. Comparison of measured and predicted RT30 in corridor. Circles: measurement, triangles: coherent prediction; crosses: incoherent prediction. (a) RT30 spectra at different source–receiver distance; (b) RT30 along the length of the tunnel at different frequency bands.

measured data are compared with predictions with the one-parameter model in the excess attenuation spectra shown in Fig. 6. They are also compared with predictions with the assumption of perfectly hard surfaces. The best-fit values for the effective flow resistivity (σ) of the ceiling and the ground surfaces are found to be 5000 kPa s m^{-2} and 400 kPa s m^{-2}, respectively. The vertical walls along the length of the corridors and the end walls are considered perfectly hard. The scattering effects that might occur from the edge between the wall and the doors are ignored. It is noted that a high flow resistivity of 5000 kPa s m^{-2} was deduced for the ceiling. This may essentially be treated as a hard boundary. However, it is found from subsequent calculations that the use of this parameter for the flow resistivity leads to better agreement with experimental data.

The loud speaker was located 4 m from one end of the corridor, equidistant from the vertical walls (i.e., along the centerline), and 0.95 m above ground. The receiver was placed along the centerline at 2 m intervals and at a height of 1.2 m above ground. Atmospheric absorption is again included in the predictions. The absorption coefficient (α) used in the incoherent model is deduced from the plane wave reflection coefficient (R_p) with the assumption of normal incidence

$$\alpha = 1 - R_p^2. \tag{13}$$

The absorption coefficients of the ground and ceiling surfaces from 100 Hz to 4000 Hz in one-third octave bands are listed in Table II.

Both RT30 and STI were measured in one-third octave bands, with the source–receiver distance ranging from 1 m to 28 m. This can be perceived as a long-distance sound propagation with respect to the cross-sectional area. Figure 7(a) shows the RT30 spectra when the receiver was located at separations of 16 m, 24 m, and 28 m from the source. The coherent model can give a satisfactory prediction up to the maximum distance at 28 m, following the pattern of the fluctuations of the measured values in most cases. On the other hand, the incoherent model can only give a general trend of the RT30. When the source and receiver were located 16 m apart, the incoherent model gives an overestimation of 0.2 s at 160 Hz, and underestimation of about 0.2 s at 1600 Hz. At source–receiver distance of 24 m, discrepancies between the incoherent predictions and measurements are found in the

TABLE II. Absorption coefficients of the ground and ceiling surfaces in the corridor.

Frequency (Hz)	100	125	160	200	250	315	400	500	630
Absorption coefficient (ceiling)	0.009	0.011	0.013	0.016	0.018	0.022	0.026	0.030	0.036
Absorption coefficient (ground)	0.058	0.068	0.081	0.094	0.110	0.130	0.152	0.177	0.206
Frequency (Hz)	800	1000	1250	1600	2000	2500	3150	4000	
Absorption coefficient (ceiling)	0.042	0.050	0.058	0.069	0.081	0.094	0.111	0.131	
Absorption coefficient (ground)	0.240	0.277	0.317	0.367	0.415	0.467	0.522	0.581	

lower frequency region, while the predictions above 1000 Hz seem agreeable. However, at a distance of 28 m, it is noticed that the coherent model generally gives a much more accurate prediction than the incoherent one. One of the main objectives of the current study is focused on the prediction of sound propagation at larger distances in a long enclosure. Although experimental measurements were also conducted at shorter distances, these measured spectra are not presented here for brevity. Nevertheless, we have displayed some of the measured results at shorter distances as shown in Fig. 7(b) for the source frequency of 250, 500, and 2500 Hz.

The variations of RT30 along the length of the corridor at frequency bands of 250 Hz, 500 Hz, and 1000 Hz are shown in Fig. 7(b). Similar to Fig. 7(a), the coherent model gives a satisfactory prediction in most cases. The coherent model is more preferable to the incoherent model, especially in the lower frequency region, where the interference effects are believed to be more significant. For the incoherent predictions, the shape of the curve does not seem to change much at the three different frequency bands, and its amplitude probably depends on the value of the absorption coefficients. The selection of absorption coefficients, or the determination of the impedance of the boundaries, plays an important role in the incoherent model. However, since normal incidence is assumed in the incoherent model, the absorption coefficients cannot represent the true physical properties of the boundary surfaces. This is even more obvious when a receiver is set further apart from the source in a long enclosure. In the coherent model, the spherical wave reflection coefficient is dependent on frequency and the angle of incidence, which is more reasonable in practice.

Results similar to those shown in Fig. 7 are obtained at other source–receiver separations and frequency bands. They are not included here for brevity. For the prediction of RT30, the average error of the coherent model was 0.05 s, with a maximum of 0.06 s at 630 Hz. On the other hand, the predicted RT30 according to the incoherent model has a higher average error of 0.07 s. Though the coherent model gives a reasonably good approximation in most cases, some discrepancies between the experimental data and coherent predictions are still found. Some possible factors include the scattering of sound from the edges between the vertical side walls and the doors and from the perforated metallic panels of the ceiling, the use of a loud speaker as a point source, and the assumption of perfectly hard vertical boundaries and end walls.

Similar to the case in the Western Harbor Tunnel, the coherent model is able to predict the STI more accurately than the incoherent model, as shown in Fig. 8. The average errors of the coherent and incoherent models were 0.03 and 0.06, respectively. There are underestimations by the incoherent model at frequency bands of 250 Hz, 500 Hz, 100 Hz, and 2000 Hz, and overestimations of at some distances at 4000 Hz. Dips and peaks are observed at 250 Hz, and the coherent model can predict the pattern well. The incoherent model is not capable of showing this interference effect. Comparing the results in Figs. 5 and 8, the advantage of the coherent model is more obvious here than in the case of the real tunnel. As mentioned before, for the coherent model, the

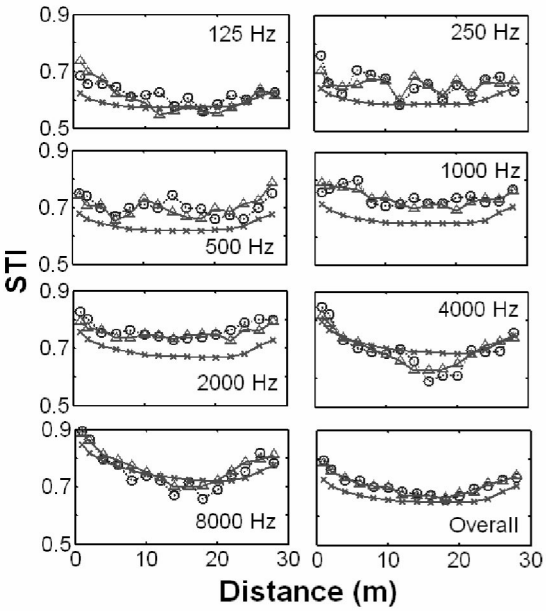


FIG. 8. Comparison of measured and predicted STI in corridor. Circles: measurement; triangles: coherent prediction; crosses: incoherent prediction.

impedance of the boundaries is characterized by the one-parameter model. It can simulate the reflection properties of the boundaries better than simply giving a nominal value as the absorption coefficient used in the incoherent model.

C. Model tunnel: Long enclosure with acoustically soft surface

In the previous section, two impedance boundaries were introduced into the model. The approximation of the coherent model is shown to be satisfactory. However, the two boundaries are still relatively hard. A third experiment was conducted to study the performance of the coherent model in long enclosures with an acoustically soft boundary. A 4.8-m-long model corridor made of hard plywood was set up in the anechoic chamber. The cross-sectional area was 0.8×1.2 m^2. The lower horizontal plane was covered with 3-cm-thick fiberglass. When the model was viewed upside-down, it simulated the situation where absorptive material is placed on the ceiling in a long enclosure where sound absorption treatment is usually made. The impedance of this boundary was calculated according to the one-parameter model with a hard-back layer.[22] This impedance model has led to a good prediction of the excess attenuation, as shown in Fig. 9(a). In the acoustic characterization of the boundary, the source and receiver were separated by a horizontal distance of 0.8 m and they were placed at a height of 0.13 m above a layer of fibreglass of 3 cm thick. The best-fit parametric value for the effective flow resistivity was 40 kPa s m^{-2}. The absorption coefficient of the fiberglass used in the incoherent model was also measured with the impedance tube. The results are presented in Fig. 9(b).

The cross section of the model tunnel was relatively smaller than those in the previous experiments, so a Tannoy driver fitted with a tube of length 1 m was used as a point source. The same microphone as mentioned before was used. The source and receiver had been placed at several different

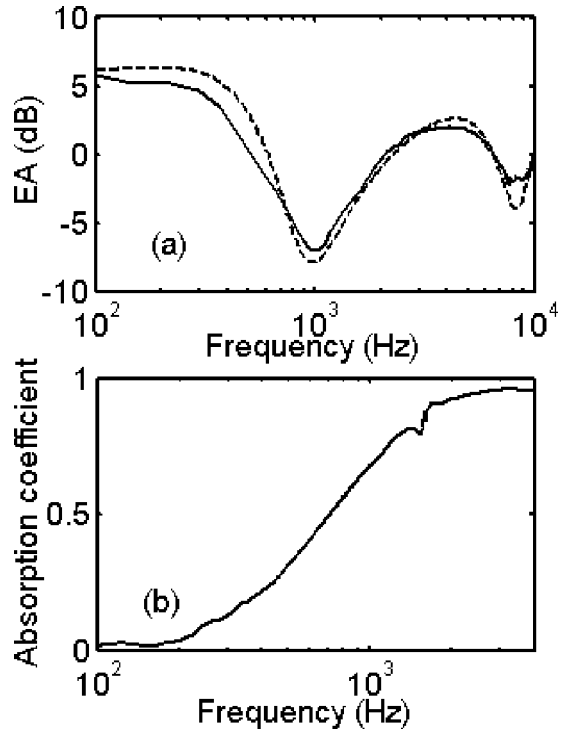


FIG. 9. Characterization of impedance of fibreglass. (a) Plot of excess attenuation against frequency. The solid line is the measurement, and the dashed line is the prediction. (b) Plot of absorption coefficient against frequency.

positions to measure the corresponding EDT and STI. Comparisons between some of the measured data and the coherent predictions are presented in Fig. 10. The source was placed 0.2 m from one of the vertical planes, and 0.13 m above the lower horizontal surface. The receiver was placed in alignment with the source along the length of the enclosure, but at a height of 0.7 m. In the same way, when the model was viewed upside down and properly scaled, it simulated the scenario of a loudspeaker being fixed near the ceiling and the wall, and a listener standing on the ground.

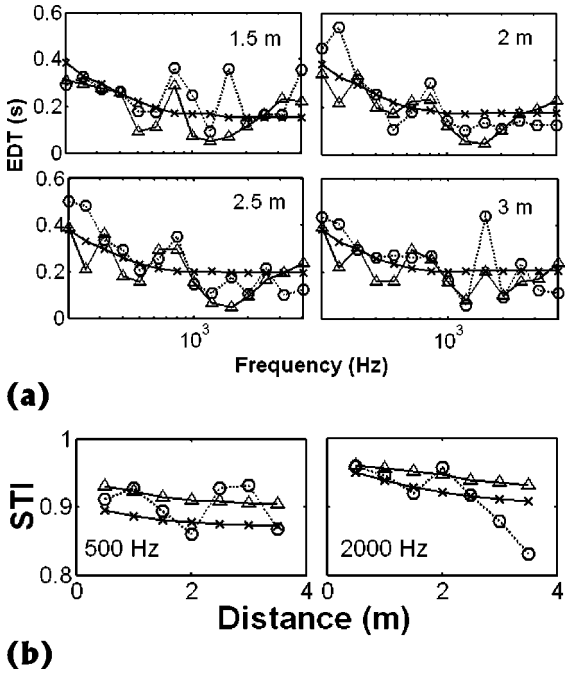


FIG. 10. Comparison of measured and predicted results in model corridor. Circles: measurement; triangles: coherent prediction; crosses: incoherent prediction. (a) EDT spectra at different source–receiver distance; (b) STI at 500 Hz and 2000 Hz octave bands.

Figure 10(a) shows the measured and predicted EDT spectrum at four different source–receiver separations, namely, 1.5 m, 2 m, 2.5 m, and 3 m. Fluctuations caused by interference effects are again observed. For example, at a distance of 1.5 m, there is a significant peak at 800 Hz, and a dip at 500 Hz. The coherent model can give a satisfactory prediction, especially between 200 Hz and 400 Hz. The predictions at other distances are also reasonably good. At a distance of 3 m, a downward slope is found from 630 Hz to 1250 Hz, which was predicted precisely by the coherent model. However, this trend is not observed in the incoherent prediction. Similar results have been obtained in other source–receiver configurations. The average error of the coherent prediction was 0.09 s, with a maximum of 0.19 s at 200 Hz.

Figure 10(b) shows the measured and predicted STI at 500 Hz and 2000 Hz octave bands. It seems that both the coherent and incoherent models give good approximations of the STI. The prediction also matches the experimental data well at other octave bands, which are not presented here for brevity. The advantage of the coherent model over the incoherent one cannot be observed clearly here. It may be due to the length of the model corridor being limited by the size of the anechoic chamber, and the source-receiver distance being relatively small. There is an overprediction at 3.5 m source–receiver separation 2000 Hz. This is probably caused by the scattering effects which occur at the edges of the model. This is also believed to be the major reason for discrepancies between the measurement results and the predictions both in EDT and STI.

IV. APPLICATION

The proposed coherent model provides a better alternative for the prediction model of reverberation time and speech transmission index, which are the governing factors of speech intelligibility in an enclosure. A design target of speech intelligibility in terms of RT and STI in long enclosures, such as train stations and long corridors, can be achieved by different methods with the help of the proposed coherent model.

In the experimental validation of the model, the signal-to-noise ratio was over 30 dB. The term in square brackets of Eq. (8) approaches 1. This means the effect of interfering noise on the prediction of STI was negligible. Therefore, Eq. (9) has been used in the prediction in the previous section. The overall STI is the weighted average of the values at the seven frequency bands. However, the signal-to-noise ratio is not always high in reality. Either the source is not powerful enough, or the noise level is high compared to the source. When the signal-to-noise ratio is small, it must be included in the prediction.

The effect of noise on the prediction of STI will be discussed with an example. When noise is excluded from the model, the coherent prediction of STI in the corridor with

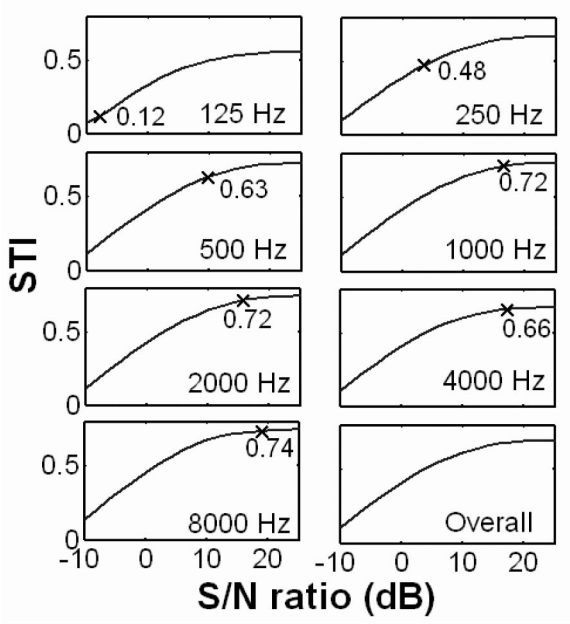


FIG. 11. Plot of STI against S/N ratio. Source–receiver distance=10 m. Cross (×): STI value at each octave band determined by the corresponding S/N ratio.

Eq. (9) is 0.7 at a source–receiver distance of 10 m. If the signal-to-noise ratio is small, Eq. (8) is then used to predict STI at the seven octave bands. The S/N ratios do not always share the same value at the seven octave bands in reality. For example, the noise produced by people talking and that by a running train in an underground train station belong to different frequency bands, and they have different noise levels. The determination of STI when the S/N ratios are different at the seven octave bands will be described here. The background noise level was taken in the corridor mentioned before. The major contribution of the noise came from the air-conditioning system, which is typical in offices and classrooms. When the signal level was turned down, the S/N ratios from 125 to 8000 Hz in octave bands were −7.8 dB, 3.4 dB, 9.7 dB, 16.4 dB, 15.7 dB, 17.3 dB, and 18.9 dB, respectively. The octave band STI can be found from the first seven curves in Fig. 11 according to the S/N ratios. They are represented by crosses in the figure. The seven values are then weighted and averaged to give a final STI, which is 0.59 in this case.

If the S/N ratios at the seven octave bands are the same, the curve of the overall STI versus S/N ratio can be plotted, as shown in Fig. 11. It increases with the S/N ratio. When the S/N ratio is high, e.g. over 20 dB, the STI value becomes steady. On the other hand, when the S/N ratio equals zero, the overall STI is just above 0.4, indicating poor speech intelligibility. One can easily identify the expected speech intelligibility at different noise conditions from the curves in Fig. 11. This provides essential information for consideration of possible acoustic treatments with respect to real-life conditions. It is also demonstrated that an accurate prediction of STI at the octave bands is crucial when the S/N ratios at the seven frequency bands are not the same in reality.

V. CONCLUSIONS

The coherent model has been extended to evaluate the impulse response of sound in a long enclosure, upon which the reverberation time and speech transmission index can be obtained. The model was validated by comparing the theoretical predictions with the experimental data in a real tunnel with hard boundaries, a long corridor with two impedance boundaries, and a model corridor with an acoustically soft boundary. It has been shown that the coherent model can give a more accurate prediction than the incoherent model, because the interference effects between contributory rays are included. The coherent model can give satisfactory predictions of both the reverberation time and the speech transmission index at different receiver positions as shown in the figures.

As mentioned in Sec. II, the impulse response is obtained by inverse Fourier transform of the predicted sound field. The number of predicted data in the single frequency will increase from low to high frequency bands, which means the computational time will also increase. The advantage of the coherent model can also be observed better in narrow band analysis, since the number of frequencies included in each band increases with the width of the band, and the interference effects might be averaged out and become less significant. Considering the computational time and the significance of the interference effects, the coherent model is therefore preferable in narrow band analysis, and in low frequency regions. It has also been demonstrated that the spherical wave reflection coefficient should not be replaced by the plane wave reflection coefficient (as assumed in many previous models), especially in cases of long source–receiver distance and enclosures with relatively soft boundaries, and in low frequency ranges.

ACKNOWLEDGMENTS

This research was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region and the Research Committee of The Hong Kong Polytechnic University. The Western Harbour Tunnel Co. Ltd. is acknowledged for granting access to its tunnel. The authors are grateful to William Fung, Mike Law, S. T. So, and T. L. Yip for their help in conducting the experiments.

[1] J. Kang, "The unsuitability of the classic room acoustical theory in long enclosures," Architectural Science Review **39**, 89–94 (1996).
[2] J. Kang, "Reverberation in rectangular long enclosures with geometrically reflecting boundaries," Acust. Acta Acust. **82**, 509–516 (1996).
[3] L. Yang and B. M. Shield, "The prediction of speech intelligibility in underground stations of rectangular cross section," J. Acoust. Soc. Am. **109**(1), 266–273 (2001).
[4] J. Kang, "Reverberation in rectangular long enclosures with diffusely reflecting boundaries," Acust. Acta Acust. **88**, 77–87 (2002).
[5] J. Kang, "A method for predicting acoustic indices in long enclosures," Appl. Acoust. **51**(2), 169–180 (1997).
[6] J. António, L. Godinho, and A. Tadeu, "Reverberation times obtained using a numerical model versus those given by simplified formulas and measurements," Acust. Acta Acust. **88**, 252–261 (2002).
[7] J. S. Suh and P. A. Nelson, "Measurement of transient response of rooms and comparison with geometrical acoustic models," J. Acoust. Soc. Am. **105**(4), 2304–2317 (1999).
[8] S. M. Dance, J. P. Roberts, and B. M. Shield, "Computer prediction of

sound distribution in enclosed spaces using an interference pressure model,'' Appl. Acoust. **44**, 53–65 (1995).
[9] K. K. Iu and K. M. Li, ''The propagation of sound in narrow street canyons,'' J. Acoust. Soc. Am. **112**(2), 537–550 (2002).
[10] K. K. Iu, ''The prediction of noise propagation in street canyons and tunnels,'' M.Phil. thesis, Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong (2003).
[11] K. M. Li and K. K. Iu, ''Propagation of sound in long enclosures,'' J. Acoust. Soc. Am. **116**, 2759–2770 (2004).
[12] K. M. Li and K. K. Iu, ''Full-scale measurements for noise transmission in tunnels,'' J. Acoust. Soc. Am. **117**, 1138–1145 (2005).
[13] T. F. W. Embleton, J. E. Piercy, and N. Olson, ''Outdoor sound propagation over ground of finite impedance,'' J. Acoust. Soc. Am. **59**(2), 267–277 (1976).
[14] M. R. Schroeder, ''New method of measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412 (1965).
[15] T. Houtgast and H. J. M. Steeneken, ''The modulation transfer function in room acoustics as a predictor of speech intelligibility,'' Acustica **28**, 66–73 (1973).
[16] H. J. M. Steeneken and T. Houtgast, ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **67**(1), 318–326 (1980).
[17] J. Kang, *Acoustics of Long Spaces* (Thomas Telford, London, 2002).
[18] M. E. Delany and E. N. Bazley, ''Acoustical properties of fibrous absorbent materials,'' Appl. Acoust. **3**, 105–116 (1970).
[19] K. Attenborough, ''Review of ground effects on outdoor sound propagation from continuous broadband sources,'' Appl. Acoust. **24**, 289–319 (1988).
[20] D. D. Rife and J. Van der Kooy, ''Transfer-function measurement with maximum-length sequences,'' J. Audio Eng. Soc. **37**, 419–443 (1989).
[21] ANSI S1.26, ''Method for calculation of the absorption of sound by the atmosphere'' (1995).
[22] K. M. Li, T. Waters-Fuller, and K. Attenborough, ''Sound propagation from a point source over extended-reaction ground,'' J. Acoust. Soc. Am. **104**, 679–685 (1998).

The influence of finite cavities on the sound insulation of double-plate structures

Jonas Brunskog[a)]
Division of Engineering Acoustics, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden

(Received 11 October 2004; revised 9 March 2005; accepted 14 March 2005)

Lightweight walls are often designed as frameworks of studs with plates on each side—a double-plate structure. The studs constitute boundaries for the cavities, thereby both affecting the sound transmission directly by short-circuiting the plates, and indirectly by disturbing the sound field between the plates. The paper presents a deterministic prediction model for airborne sound insulation including both effects of the studs. A spatial transform technique is used, taking advantage of the periodicity. The acoustic field inside the cavities is expanded by means of cosine-series. The transmission coefficient (angle-dependent and diffuse) and transmission loss are studied. Numerical examples are presented and comparisons with measurement are performed. The result indicates that a reasonably good agreement between theory and measurement can be achieved. [DOI: 10.1121/1.1904264]

PACS numbers: 43.55.Ti, 43.40.Dx, 43.20.Tb, 43.55.Rg [NX] Pages: 3727–3739

I. INTRODUCTION

New building systems are often designed as lightweight systems. One advantage of these systems is the possibility they provide of lowering house production costs. The interest in such lightweight building techniques is therefore large. Acoustically there are both advantages and disadvantages with such structures. One of the advantages is that a wall consisting of two plates—a double-plate structure—provides good insulation against airborne noise in relation to its weight. There is, however, no theoretical prediction model for sound-insulation that takes into account all important aspects of such a wall system.

If a double-plate structure consists of a framework of studs, the studs will not only influence the vibration field directly, i.e., short-circuiting the plates as sound bridges,[1] but also affect the acoustic field in the cavities. The studs can be seen as walls in the cavities, thus introducing finiteness, which leads to resonance. Consider therefore a double-leaf structure excited by an incoming wave on the source side. The plate on the source side is excited and will radiate to the cavities and excite the framing beams. The plate on the receiver side is then excited by the acoustic field in the cavities and by the vibration of the beams, and will radiate to the surrounding acoustic fields.

In building acoustics it is common to use Statistical Energy Analysis (SEA), different energy-flow methods such as those associated with the new European standard for computing building acoustics [EN 12354 (Ref. 2)], or various semiempirical methods to find prediction models. Such approaches are often preferable when details of a standard type of construction are considered, for example, when well-known elements are combined in EN 12354, or when large variations in the material or geometric data affect the results very little. An approach of this type is basically pragmatic, emphasis being placed on achieving reasonable results quickly. The amount of information used to account for the physics involved is minimal, each building element being described by a number (for each frequency); the mean of the sound energy being obtained for each part or component of the structure separately. The fact that only a minimal amount of information is utilized here represents both an advantage and a disadvantage. These approaches are particularly appropriate for dealing with homogeneous and clearly distinguishable building elements, such as in traditional building construction systems in which the elements are heavy and homogeneous. Such approaches are not likely to be successful, however, if one's interest is in discovering new types of solutions, since the lack of the information needed makes it impossible to describe the physics of the situation adequately. Moreover, wood frame building elements (or other stud-plate building elements) may not be considered as homogeneous and isotropic. Hence, they do not meet the basic requirements for simplified prediction models (such as EN and basic SEA models) where it is assumed that all building elements can be considered to be a single subsystem. This has also been observed in measurements.[3] For studies of SEA applied to sound insulation with double wall systems, Craik[4] and Craik and Smith[5] are examples. As an example of semiempirical prediction models, the model described by Sharp[6] should be mentioned.

Instead, an analytical/deterministic approach is used in the present paper. The approach is largely based on the following papers. The classical work on double-leaf walls is made by London.[7] This work, however, does not take into account the studs or the finiteness of the cavities. Cremer *et al.*[1] describe two parallel plates connected via a locally reacting stiffness layer. A recent similar study by Kropp and Rebillard,[8] treats airborne sound insulation of double-wall constructions without studs. The focus is on the possibility of optimizing the sound insulation. Lin and Garrelick[9] investigated the transmission of a plane wave through two infinite parallel plates connected by periodic studs that behave as rigid bodies. A fluid coupling in the cavities between the

a)Electronic mail: jonas.brunskog@acoustics.lth.se

plates was also present, which let the airborne waves pass unaffected through the beams. The two systems were solved simultaneously by means of Fourier transforms and periodic considerations. Recent measurements on wood stud walls, made by Bradley and Birta,[10] show that the Lin and Garrelick theory explains the most important low-frequency features of sound transmission through these wood stud walls. Rumerman[11] and Mace[12] treated a single infinite plate reinforced by beams and excited by a convected pressure (that could be a incoming plane wave) using Fourier transforms. Takahashi[13] considered noise control in buildings having double-plate walls. Each structure considered consisted of two parallel plates of infinite extent connected by various connectors. The connectors were point connectors or rib-stiffeners, and the sound field in the cavity passed unaffected through them. The structures were driven by point forces, and the resulting sound radiation was studied. The problem of acoustic reflection of a double-plate system with periodic connectors were studied by Skelton,[14] who also assumed the connectors to be invisible to the fluid in the cavity. The connectors were applied in form of longitudinal plate waveguides. The related problem of acoustic reflection of a plate with periodic ribs and a back cavity was studied by Sakagami *et al.*[15] Also here the ribs were acoustically transparent so that the sound field in the back cavity was not influenced by them. The only paper found that takes account of the influence of ribs or connectors on the cavity field is Skelton,[16] who considered the effect of a single connector, rigid with respect to the fluid, using the Wiener–Hopf technique. Also Urusovskii[17] studied a periodic double plate system, but used a space-harmonic assumption and truncated the infinite system of equation that arose. The cavity field was unaffected by the beams. The present author and Hammer have studied impact sound transmissions in lightweight floors using transform technique,[18] including the beams and a cavity, but letting the waves go unaffected through the beams (as in Refs. 9, 13, 14, 17, and 15). Maysenhoelder and Haberkern[19] studied sound transmission through two-dimensionally periodic inhomogeneous infinite plates using a three-dimensional elastic description of the wall (which typically is a masonry wall). A periodic assumption similar to the space-harmonic assumption was used together with the Hamilton principle. The problem under focus in the present paper can in principle also be treated with their method. A literature survey of available deterministic prediction methods that takes periodicity into account can be found in Ref. 20 (with focus on point excitation).

The approach in the present paper is similar to the one introduced by Rumerman[11] and Mace,[12] but the treatment of the cavities is original for the present paper. An expansion in a suitable orthogonal series is made for the cavity field in order to take into account the finiteness of the cavities. The solution of the displacement field will then consist of one term originating from the excitation, terms originating from the structural coupling via the beams, and terms originating from the fluid coupling via the cavity. The former two are equivalent to the existing theory and the latter are new terms. The paper is comprised as follows: Following the Introduction, the problem under consideration is formulated. In Sec. III, the cavity reactions and field are studied, and the remaining reaction forces are collected in Sec. IV. In Sec. V, the solution is described. Sound radiation and sound insulation are considered in Sec. VI. The numerical results are presented and discussed in Sec. VII, and the paper ends with some concluding remarks.

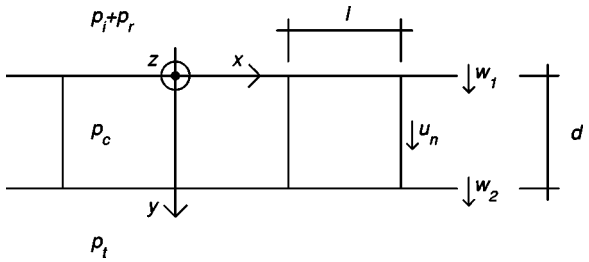


FIG. 1. A double leaf structure. The horizontal lines are the center lines of the plates with displacements w_1 and w_2. The vertical lines are the center lines of the beams with displacements u_n.

II. FORMULATION OF THE PROBLEM

Consider a double-leaf wall stiffened with studs, as in Fig. 1. The studs are assumed to be infinitely stiff in bending round the *z*-axis and of zero thickness; the beams are motionless in the *x*-direction. However, they are allowed to bend round the *x*-axis, yielding flexural displacement u_n in the *y*-direction and bending wave propagation in the *z*-direction. The structure is infinite in both the *x*- and *z*-direction. The studs are located periodically with equal distance l between them. The studs are assumed to not cause any moment reactions or coupling; the connection plate-stud is assumed to be in form of a pin-joint. The displacement of the plates are denoted w_1 and w_2, and the displacement of the *n*th beam is denoted u_n.

In the field in front of the first plate a pressure p_i due to an incident wave, and a reaction pressure p_r due to radiated waves is present. In the field backing the second plate a reaction pressure p_t is present due to transmitted waves. The incident pressure is of the form

$$p_i = \hat{p}_i \mathrm{e}^{-\mathrm{i}(k_x x + k_y y + k_z z - \omega t)} \tag{1}$$

where a possible choice of the wave numbers is

$$k_x = k \sin\theta \cos\varphi, \quad k_z = k \sin\theta \sin\varphi, \quad k_y = k \cos\theta,$$

i.e., an incoming wave with wave number $k = \omega/c$. θ is the angle between the *y*-axis and the direction of the incident wave and φ is the azimuth angle. The time dependence and the *z*-dependence, $\exp(\mathrm{i}\omega t - \mathrm{i}k_z z)$, will henceforth be suppressed throughout. Further, since the structure is periodic in *x* the response satisfies the periodicity relation, see e.g., Ref. 12 or Ref. 20

$$w_i(x+l) = w_i(x)\mathrm{e}^{-\mathrm{i}k_x l} \tag{2}$$

usually denoted as Floquet's principle.

A. Governing equations

Consider the plates of the double-leaf structure. The system of governing equations that is to be solved can be written as

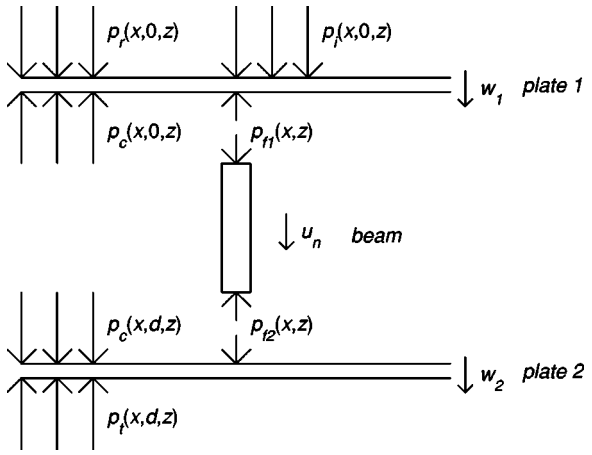


FIG. 2. Model for the equation of motion; definition of pressures acting on the plates and the beams in the vertical direction.

$$D_1'\nabla^4 w_1 - m_1''\omega^2 w_1 = p_i|_{y=0} - p_c|_{y=0} + p_r|_{y=0} - p_{f1}, \tag{3}$$

$$D_2'\nabla^4 w_2 - m_2''\omega^2 w_2 = p_c|_{y=d} - p_t|_{y=d} + p_{f2} \tag{4}$$

for the first and the second plate, respectively, where

$$\nabla^4 = \frac{\partial^4}{\partial x^4} - 2k_z^2\frac{\partial^2}{\partial x^2} + k_z^4$$

and D_i' and m_i'' are the flexural rigidity and mass per unit area of plate number i, respectively, $\omega = 2\pi f$ is the angular frequency and d the distance between the plates. It is convenient to decompose the reflected pressure into two components, $p_r = p_g + p_s$, where p_g is the reflected pressure generated by a rigid reflector (geometrical reflection), so that

$$p_g|_{y=0} = p_i|_{y=0}, \tag{5}$$

and p_s is the scattered part due to the elastic motion of the structure. Figure 2 shows the different reaction fields. The reaction pressures from the surrounding fluid can be assumed to be coupled to the displacements field by operators,

$$p_s|_{y=0} = \mathcal{R}w_1, \tag{6}$$

$$p_t|_{y=d} = \mathcal{T}w_2, \tag{7}$$

where $\mathcal{R}$ and $\mathcal{T}$ are linear operators that will be determined in Sec. IV. The cavity pressure p_c and frame pressure p_f are treated in separate sections, due to their complexity.

B. Transformed equations

The Fourier transform of the displacement w_i with respect to the coordinate x and the corresponding inverse transform is defined as

$$\tilde{w}_i(\alpha) = \int_{-\infty}^{\infty} w_i(x)\mathrm{e}^{\mathrm{i}\alpha x}\mathrm{d}x, \tag{8}$$

$$w_i(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{w}_i(\alpha)\mathrm{e}^{-\mathrm{i}\alpha x}\mathrm{d}\alpha. \tag{9}$$

Thus, the Fourier transform over x of the incoming wave yields a Dirac function, $\tilde{p}_i = 2\pi\hat{p}_i\delta(\alpha - k_x)$. For the reaction pressure, the transform yields algebraic expressions instead of operators. Thus, the transformed pressures are

$$(\tilde{p}_i + \tilde{p}_g)|_{y=0} = 4\pi\hat{p}_i\delta(\alpha - k_x),$$

$$\tilde{p}_s|_{y=0} = R\tilde{w}_1,$$

$$\tilde{p}_t|_{y=d} = T\tilde{w}_2,$$

where R and T are the transformed version of the operators $\mathcal{R}$ and $\mathcal{T}$. Applying the Fourier transform with respect to x to Eqs. (3) and (4), taking into account Eqs. (6) and (7), gives

$$\begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}\begin{bmatrix} \tilde{w}_1 \\ -\tilde{w}_2 \end{bmatrix} = 4\pi\begin{bmatrix} \hat{p}_i \\ 0 \end{bmatrix}\delta(\alpha - k_x) - \begin{bmatrix} \tilde{p}_{f1} \\ \tilde{p}_{f2} \end{bmatrix} - \begin{bmatrix} \tilde{p}_c|_o \\ \tilde{p}_c|_d \end{bmatrix}, \tag{10}$$

where

$$S_1(\alpha) = D_1'(\alpha^2 + k_z^2)^2 - m_1''\omega^2 - R(\alpha), \tag{11}$$

$$S_2(\alpha) = D_2'(\alpha^2 + k_z^2)^2 - m_2''\omega^2 + T(\alpha), \tag{12}$$

are spatial stiffnesses. The solution of Eq. (10) and its inverse transform is given in Sec. V. It is, however, first necessary to give expressions for the reaction pressures.

III. THE CAVITY

The reaction pressures due to the cavity field are examined in this section.

A. The cavity, a parallelepipedic space

Consider again Fig. 1, where a fluid is occupying the space $0 < y < d$, divided into subspaces $ml \leqslant x \leqslant (m+1)l$, $m = 0, \pm 1, \pm 2, ..., \pm\infty$. An acoustic pressure $p_c(x,y)$ is present in the fluid. The acoustic pressure satisfies the Helmholtz equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)p_c + (k_c^2 - k_z^2)p_c = 0, \tag{13}$$

where $k_c = \omega/c_c$ is the wavenumber in the cavity (possibly different from k, the wave number in the surrounding fluid) and c_c is the speed of sound in the medium. The acoustic pressure also satisfies the boundary conditions

$$\left[\frac{\partial p_c}{\partial y}\right]_{y=0} = \omega^2\rho_c w_1, \tag{14}$$

$$\left[\frac{\partial p_c}{\partial y}\right]_{y=d} = \omega^2\rho_c w_2, \tag{15}$$

where ρ_c is the density of the fluid. Equations (14) and (15) ensure equality of the fluid velocity at the plate surface and the plate velocity. These boundary conditions are the ones that are fulfilled in the papers.[9,13,16–18] In the present paper a new set of boundary conditions is also to be fulfilled, as a consequence of the studs being nonmovable in the x-direction, namely

$$\left[\frac{\partial p_c}{\partial x}\right]_{x=ml} = 0, \quad m = 0, \pm 1, \pm 2, ..., \pm\infty \tag{16}$$

ensuring absence of displacement in the x-direction at the "rigid walls" of the studs at $x = ml$. Divide the field into subfields corresponding to the cavities

$$p_c(x,y)=\sum_{m=-\infty}^{\infty} p_c^{(m)}(x,y)\Theta(x,ml,ml+l), \tag{17}$$

where

$$\Theta(x,ml,ml+l)\equiv\theta(x-ml)-\theta(x-(m+1)l)$$

and where $\theta(x)$ is the Heavisides step function and $\Theta(x,a,b)$ is the ''hat'' function that equals unity between a and b and is zero otherwise. Assume that the pressure field in the mth cavity can be written as an orthogonal sum of cosine functions in the x-direction. Define ε_n,

$$\varepsilon_n\equiv\begin{cases}\frac{1}{2} & \text{if } n=0\\ 1 & \text{if } n\neq 0.\end{cases}$$

Thus, the cosine expansion can be written

$$p_c^{(m)}(x,y)=\sum_{n=0}^{\infty}\varepsilon_n p_{c,n}^{(m)}(y)\cos(n\pi x/l), \tag{18}$$

where $ml\leqslant x\leqslant(m+1)l$. It is easily shown that this assumption fulfills the boundary conditions (16) directly, as well as the Helmholtz equation (13). The boundary conditions in (14) and (15) is still to be fulfilled by means of specifying the yet unknown function $p_{c,n}^{(m)}(y)$, see Sec. III B. When this is achieved, as both differential equation and boundary condition then are fulfilled, the solution to the cavity field is found.

The periodicity, expressed in Eq. (2), is now to be taken into account. This implies that the pressure acting on two neighboring bays are related to each other through a phase difference

$$p_c^{(m+1)}=p_c^{(m)}\mathrm{e}^{-\mathrm{i}k_x l}$$

and especially

$$p_c^{(m)}=p_c^{(0)}\mathrm{e}^{-\mathrm{i}k_x ml}.$$

Hence, Eq. (17) reduces to

$$p_c(x,y)=p_c^{(0)}(x,y)\sum_{m=-\infty}^{\infty}\Theta(x,ml,ml+l)\mathrm{e}^{-\mathrm{i}k_x ml}. \tag{19}$$

Thus, the total field in the cavities is determined by the field in the 0th-cavity. Expressing this field in terms of a cosine series, Eq. (18), and inserting in Eq. (19), yields

$$p_c(x,y)=\sum_{n=0}^{\infty}\varepsilon_n p_{c,n}^{(0)}(y)\cos(n\pi x/l)\times\sum_{m=-\infty}^{\infty}\Theta(x,ml,ml+l)\mathrm{e}^{-\mathrm{i}k_x ml}. \tag{20}$$

Hence, the two sums are separated. The spatial Fourier transform of the cavity reaction pressure formally is

$$\tilde{p}_c(\alpha,y)=\int_{-\infty}^{\infty}\sum_{n=0}^{\infty}\varepsilon_n p_{c,n}^{(0)}\cos(n\pi x/l)\times\sum_{m=-\infty}^{\infty}\Theta(x,ml,ml+l)\mathrm{e}^{-\mathrm{i}k_x ml}\mathrm{e}^{\mathrm{i}\alpha x}\mathrm{d}x \tag{21}$$

and will be further treated in Sec. III D.

B. The field in the 0th cavity

The cosine expansion (18) of the cavity field is inserted into the Helmholtz equation (13). It can be shown that this expression reduces to

$$\frac{\partial^2 p_{c,n}^{(0)}(y)}{\partial y^2}+\left(k_c^2-\left(\frac{n\pi}{l}\right)^2-k_z^2\right)p_{c,n}^{(0)}(y)=0,\quad n=0,1,2,\ldots. \tag{22}$$

Define a propagation number for the nth component

$$\gamma_n^2\equiv\left(\frac{n\pi}{l}\right)^2+k_z^2-k_c^2.$$

The solution can be written as one wave in the positive y-direction and one in the negative y-direction,

$$p_{c,n}^{(0)}(y)=\hat{p}_{c,n+}\mathrm{e}^{-\gamma_n y}+\hat{p}_{c,n-}\mathrm{e}^{\gamma_n y} \tag{23}$$

and one differentiation with respect to y gives

$$\frac{\partial p_{c,n}^{(0)}(y)}{\partial y}=\gamma_n(-\hat{p}_{c,n+}\mathrm{e}^{-\gamma_n y}+\hat{p}_{c,n-}\mathrm{e}^{\gamma_n y}). \tag{24}$$

In order to determine $p_{c,n\pm}$, and thereby $p_{c,n}^{(0)}(y)$, the remaining boundaries are now expressed in the same cosine series as the cavity field

$$w_i(x)=\sum_{n=0}^{\infty}\varepsilon_n w_{i,n}\cos(n\pi x/l), \tag{25}$$

for $0\leqslant x\leqslant l$, where

$$w_{i,n}=\frac{2}{l}\int_0^l w_i(x)\cos(n\pi x/l)\mathrm{d}x. \tag{26}$$

It should here be noted that Eqs. (25) and (26) will not put any constraint to the solution of w_i; it is just a description of the displacement field in terms of the cosine basis used in the cavity. Also note that the structural field is given in the form of Eqs. (52) and (53) and not in the form of Eq. (25). The boundary conditions (14)–(15) have to be fulfilled by every Fourier component, connecting $p_{c,n}^{(0)}$ to $w_{i,n}$ [as seen if (18) and (25) is inserted into the boundary condition],

$$\left.\frac{\partial p_{c,n}^{(0)}}{\partial y}\right|_{y=0}=\omega^2\rho_c w_{1,n},\quad \left.\frac{\partial p_{c,n}^{(0)}}{\partial y}\right|_{y=d}=\omega^2\rho_c w_{2,n}.$$

Taking into account the standing wave and its derivatives (23)–(24) gives a system of equations, which yields the following amplitudes of the components in the standing wave:

$$\hat{p}_{c,n+}=\frac{\omega^2\rho_c}{\gamma_n}\frac{(w_{2,n}-w_{1,n}e^{\gamma_n d})}{2\sinh(\gamma_n d)}, \tag{27}$$

$$\hat{p}_{c,n-}=\frac{\omega^2\rho_c}{\gamma_n}\frac{(w_{2,n}-w_{1,n}e^{-\gamma_n d})}{2\sinh(\gamma_n d)}. \tag{28}$$

Insert Eqs. (27) and (28) in Eq. (23),

$$p_{c,n}^{(0)}(y)=\frac{\omega^2\rho_c}{\sinh(\gamma_n d)\gamma_n}(w_{2,n}\cosh(\gamma_n y)-w_{1,n}\cosh(\gamma_n(d-y))). \tag{29}$$

Putting $y=0$ and $y=d$, respectively, and rewriting in a matrix form yields

$$\begin{bmatrix} p_{c,n}^{(0)}(0) \\ p_{c,n}^{(0)}(d) \end{bmatrix} = \frac{-\omega^2\rho_c}{\gamma_n \sinh \gamma_n d} \begin{bmatrix} \cosh(\gamma_n d) & 1 \\ 1 & \cosh(\gamma_n d) \end{bmatrix} \times \begin{bmatrix} w_{1,n} \\ -w_{2,n} \end{bmatrix}, \tag{30}$$

which is a connection between the cavity reaction pressure and the structural displacement, expressed in terms of cosine-series components.

C. Solution of the cavity pressure field

The cavity reaction pressures used in Eqs. (3) and (4) can then be expressed as, if Eq. (30) is inserted in Eq. (20),

$$\begin{bmatrix} p_c(x,0) \\ p_c(x,d) \end{bmatrix} = \left(\sum_{n=0}^{\infty} \varepsilon_n \mathbf{J}_n \begin{bmatrix} w_{1,n} \\ -w_{2,n} \end{bmatrix} \cos(n\pi x/l) \right) \times \sum_{m=-\infty}^{\infty} \Theta(x, ml, ml+l) \mathrm{e}^{-\mathrm{i}k_x ml}, \tag{31}$$

where the following definition $\mathbf{J}_n$ of the matrix in Eq. (30) has been introduced,

$$\mathbf{J}_n \equiv \frac{-\omega^2\rho_c}{\gamma_n \sinh(\gamma_n d)} \begin{bmatrix} \cosh(\gamma_n d) & 1 \\ 1 & \cosh(\gamma_n d) \end{bmatrix}.$$

The displacement field in the plates is unknown, and needs to be solved for. The components $w_{i,n}$ are expressed as integrals in Eq. (26).

D. Fourier transform of the cavity pressure

The spatial Fourier transform of the cavity reaction pressure is formally given in Eq. (21). Define an auxiliary function

$$\Phi_n(\alpha) \equiv \mathcal{F}_x \left\{ \cos\left(\frac{n\pi x}{l}\right) \sum_{m=-\infty}^{\infty} \Theta(x, ml, ml+l) \mathrm{e}^{-\mathrm{i}k_x ml} \right\}, \tag{32}$$

where $\mathcal{F}_x\{\cdot\} = \int \cdot \exp(\mathrm{i}\alpha x)\mathrm{d}x$ is the spatial Fourier transform operator, see Eq. (8). Then, the transformed cavity pressures at the boundaries are

$$\begin{bmatrix} \tilde{p}_c(\alpha,0) \\ \tilde{p}_c(\alpha,d) \end{bmatrix} = \sum_{n=0}^{\infty} \begin{bmatrix} J_{11,n} & J_{12,n} \\ J_{21,n} & J_{22,n} \end{bmatrix} \begin{bmatrix} w_{1,n} \\ -w_{2,n} \end{bmatrix} \varepsilon_n \Phi_n(\alpha), \tag{33}$$

where $J_{11,n}$, etc. are the components of $\mathbf{J}_n$. A more suitable form of the function $\Phi_n(\alpha)$ is needed. Some manipulations, including use of the Poisson sum formula, yields

$$\Phi_n(\alpha) = \frac{\mathrm{i}\pi}{l} \frac{1-\mathrm{e}^{\mathrm{i}\alpha l}\mathrm{e}^{-\mathrm{i}n\pi}}{\alpha - n\pi/l} \sum_{m=-\infty}^{\infty} \delta(\alpha - k_x + (n-2m)\pi/l) + \frac{\mathrm{i}\pi}{l} \frac{1-\mathrm{e}^{\mathrm{i}\alpha l}\mathrm{e}^{\mathrm{i}n\pi}}{\alpha + n\pi/l} \sum_{m=-\infty}^{\infty} \delta(\alpha - k_x - (n+2m)\pi/l). \tag{34}$$

For details see Appendix A. Moreover, due to the infinite sums, we have

$$\sum_{m=-\infty}^{\infty} \delta(\alpha - k_x + (\pm n - 2m)\pi/l) = \sum_{m=-\infty}^{\infty} \delta(\alpha - k_x - 2m\pi/l).$$

Thus, we can rewrite Eq. (34),

$$\Phi_n(\alpha) = \frac{2\pi}{l} \varsigma_n(\alpha) \sum_{m=-\infty}^{\infty} \delta(\alpha - k_x - 2m\pi/l), \tag{35}$$

where

$$\varsigma_n(\alpha) \equiv \mathrm{i}\alpha \frac{1-\mathrm{e}^{\mathrm{i}\alpha l}(-1)^n}{\alpha^2 - (n\pi/l)^2}. \tag{36}$$

Combining Eqs. (33) and (35) yields the transformed cavity reaction pressures

$$\begin{bmatrix} \tilde{p}_c(\alpha,0) \\ \tilde{p}_c(\alpha,d) \end{bmatrix} = \frac{2\pi}{l} \sum_{n=0}^{\infty} \begin{bmatrix} J_{11,n} & J_{12,n} \\ J_{21,n} & J_{22,n} \end{bmatrix} \begin{bmatrix} w_{1,n} \\ -w_{2,n} \end{bmatrix} \varepsilon_n \varsigma_n(\alpha) \times \sum_{m=-\infty}^{\infty} \delta(\alpha - k_x - 2m\pi/l). \tag{37}$$

E. Displacement field in the cavity

The displacement field inside a cavity is now to be considered. It is only necessary to consider the 0th cavity due to the periodic relation (2). The pressure field in the 0th cavity is given in Eq. (18) where index $m=0$ is to be used, and $p_{c,n}^{(0)}$ is given in Eq. (29). The corresponding displacement field in the cavity is

$$w_{c,x}^{(0)} = \frac{1}{\rho_c \omega^2} \frac{\partial p_c^{(0)}}{\partial x}, \quad w_{c,y}^{(0)} = \frac{1}{\rho_c \omega^2} \frac{\partial p_c^{(0)}}{\partial y}$$

for displacement in the x- and y-direction, respectively. Derivation of Eq. (18) yields

$$w_{c,x}^{(0)} = -\sum_{n=0}^{\infty} \varepsilon_n \frac{n\pi}{l\gamma_n} \frac{\sin(n\pi x/l)}{\sinh(\gamma_n d)} (w_{2,n} \cosh(\gamma_n y) - w_{1,n} \cosh(\gamma_n (d-y))), \tag{38}$$

$$w_{c,y}^{(0)} = \sum_{n=0}^{\infty} \varepsilon_n \frac{\cos(n\pi x/l)}{\sinh(\gamma_n d)} (w_{2,n} \sinh(\gamma_n y) + w_{1,n} \sinh(\gamma_n (d-y))) \tag{39}$$

for the x- and y-direction, respectively.

IV. THE REMAINING REACTION FORCES

In order to get a complete solution, the forces and pressures involved must be specified.

A. The frame reactions

Consider Fig. 3, describing the nth beam and the corresponding reaction line forces. The continuity equation at each plate-beam connection point is assumed to be springlike

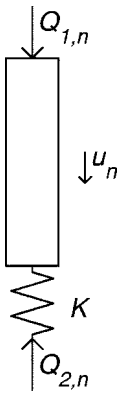


FIG. 3. The nth beam. Reaction line forces $Q_{1,n}$ and $Q_{2,n}$, displacement u_n, and line stiffness K.

and takes the following form, if the displacement $u_n(z)$ of the nth beam equals the displacement of the first plate, $u_n(z)=w_1(nl,z)$, and

$$Q_{1,n}(z)-Q_{2,n}(z)=\mathcal{G}w_1(nl,z), \tag{40}$$

$$Q_{2,n}(z)=K(w_1(nl,z)-w_2(nl,z)), \tag{41}$$

where $n=-\infty,\ldots,\infty$, and where $\mathcal{G}$ is a linear operator, K is a stiffness, and $Q_{i,n}$ are line forces. For the nth beam,

$$\begin{bmatrix} Q_{1,n}(z) \\ Q_{2,n}(z) \end{bmatrix} = \begin{bmatrix} \mathcal{G}+K & K \\ K & K \end{bmatrix} \begin{bmatrix} w_1(nl,z) \\ -w_2(nl,z) \end{bmatrix}. \tag{42}$$

The linear operator corresponding to a Euler beam is

$$\mathcal{G}=E_fI_f\frac{d^4}{dz^4}-\rho_fA_f\omega^2, \tag{43}$$

where E_fI_f is the bending rigidity and ρ_fA_f is mass per unit length of the beam. The z-derivatives will be replaced by k_z^4 due to the assumed e^{-ik_zz} dependency in Eq. (1). Thus, the algebraic expression for transformed operator G is found,

$$G=E_fI_fk_z^4-\rho_fA_f\omega^2. \tag{44}$$

The frame reaction pressure is

$$p_{fi}=\sum_{m=-\infty}^{\infty} Q_{i,m}\delta(x-ml), \tag{45}$$

for $i=1,2$. The displacement fields w_1 and w_2 satisfy the periodicity relation (2) since the structure and driving are periodic. Therefore

$$w_i(nl)=w_i(0)e^{-ink_xl}, \tag{46}$$

for $i=1,2$. Therefore, using Eqs. (40), (41), and (46), the reaction fields caused by the beams, used in Eqs. (3) and (4), can be written

$$\begin{bmatrix} p_{f1} \\ p_{f2} \end{bmatrix} = \begin{bmatrix} G+K & K \\ K & K \end{bmatrix} \begin{bmatrix} w_1(0,z) \\ -w_2(0,z) \end{bmatrix} \sum_{m=-\infty}^{\infty} e^{-imk_xl}\delta(x-ml). \tag{47}$$

Fourier transform Eq. (47), makes use of the Poisson sum formulas (A7) and (A8),

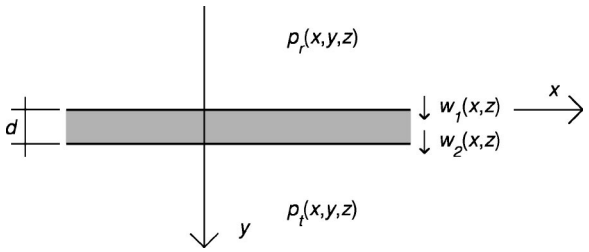


FIG. 4. The acoustic pressure.

$$\begin{bmatrix} \tilde{p}_{f1} \\ \tilde{p}_{f2} \end{bmatrix} = \frac{2\pi}{l} \begin{bmatrix} G+K & K \\ K & K \end{bmatrix} \begin{bmatrix} w_1(0,z) \\ -w_2(0,z) \end{bmatrix} \times \sum_{m=-\infty}^{\infty} \delta(\alpha-2m\pi/l-k_x). \tag{48}$$

B. The reactions from the transmitted and reflected fields

The transformed versions of the operators $\mathcal{R}$ and $\mathcal{T}$ in Eqs. (6) and (7), denoted R and T, respectively, are to be determined. Consider Fig. 4, where a fluid is occupying the upper half space with an acoustic pressure $p_s(x,y,z)$, $y\leqslant 0$, and the lower half space is occupied by a fluid with an acoustic pressure $p_t(x,y,z)$, $y\geqslant d$. The total pressure field in the incident side is $p_i+p_g+p_s$, as was discussed more closely in Sec. II. It is assumed that the two fields have the same sound speed c and density ρ. Two moving surfaces occupy the x-z-plane in $y=0$ and $y=d$, vibrating with displacements $w_1(x,z)$ and $w_2(x,z)$. The scattered and transmitted acoustic pressure field can be expressed as

$$p(x,y,z)=\begin{cases} p_s(x,y,z) & y\leqslant 0 \\ p_t(x,y,z) & y\geqslant d. \end{cases} \tag{49}$$

The acoustic pressure satisfies the Helmholtz equation

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}-k_z^2\right)p+\frac{\omega^2}{c^2}p=0, \tag{50}$$

where c is the speed of sound and ω is the radian frequency, together with the boundary conditions, the same as in (14) and (15) but with p_s in the former equation and p_t in the latter, ensuring equality of the fluid velocity at the plate surface and the plate velocity. The Helmholtz equation (50) is now transformed in the x-direction, using Eq. (8). This yields an one-dimensional Helmholtz equation in the y-direction. As only outgoing waves are of interest, the solution of the transformed equation is

$$\tilde{p}(\alpha,y)=\begin{cases} \hat{p}_se^{\gamma y}, & y<0 \\ \hat{p}_te^{-\gamma(y-d)}, & y>d, \end{cases} \tag{51}$$

where

$$\gamma^2 \equiv \alpha^2 + k_z^2 - k^2$$

and where γ has to be evaluated so that $\Re\{\gamma\} \geqslant 0$, $\Im\{\gamma\} \geqslant 0$ if $\Re\{\gamma\} = 0$. Therefore, using the boundary conditions (14) and (15) together with the derivative of (51), for $y = 0$ and $y = d$, yields the identification of the coefficients,

$$R = -T = \frac{\omega^2 \rho}{\gamma(\alpha)}.$$

V. THE INVERSE TRANSFORM AND THE SOLUTION OF THE PROBLEM

The solution to the problem can now be found by means of applying the inverse Fourier transform in Eq. (10), making use of Eqs. (37) and (48). The Dirac functions ensure that the inverse transform can be easily performed, and the displacement field can thus be determined. The displacements are

$$w_1(x) = \frac{2\hat{p}_i e^{-ik_x x}}{S_1(k_x)} - \frac{1}{l}((G+K)w_1(0) - Kw_2(0))T_1^{(f)}(x) - \frac{1}{l}\sum_{n=0}^{\infty}(J_{11,n}w_{1,n} - J_{12,n}w_{2,n})\varepsilon_n T_{1,n}^{(c)}(x) \quad (52)$$

and

$$w_2(x) = \frac{1}{l}(Kw_1(0) - Kw_2(0))T_2^{(f)}(x) + \frac{1}{l}\sum_{n=0}^{\infty}(J_{21,n}w_{1,n} - J_{22,n}w_{2,n})\varepsilon_n T_{2,n}^{(c)}(x), \quad (53)$$

where the following abbreviations have been used:

$$T_j^{(f)}(x) \equiv \sum_{m=-\infty}^{\infty} \frac{e^{-i\alpha_m x}}{S_j(\alpha_m)}, \quad (54)$$

$$T_{j,n}^{(c)}(x) \equiv \sum_{m=-\infty}^{\infty} \frac{\varsigma_n(\alpha_m)e^{-i\alpha_m x}}{S_j(\alpha_m)}, \quad (55)$$

where $\alpha_m = k_x + 2m\pi/l$ and where $j = 1,2$.

Note here that the values of $w_1(0)$, $w_2(0)$ and the Fourier components $w_{j,n}$ are still unknown. To determine the first two of them, let $x \to 0$ in Eq. (52), for the first plate

$$w_1(0) = \frac{2\hat{p}_i}{S_1(k_x)} - \frac{(G+K)}{l}T_1^{(f)}(0)w_1(0) + \frac{K}{l}T_1^{(f)} \times (0)w_2(0) - \frac{1}{l}\sum_{n=0}^{\infty} J_{11,n}\varepsilon_n T_{1,n}^{(c)}(0)w_{1,n} + \frac{1}{l}\sum_{n=0}^{\infty} J_{12,n}\varepsilon_n T_{1,n}^{(c)}(0)w_{2,n} \quad (56)$$

and Eq. (53) for the second plate

$$w_2(0) = \frac{KT_2^{(f)}(0)}{l}w_1(0) - \frac{KT_2^{(f)}(0)}{l}w_2(0) + \frac{1}{l}\sum_{n=0}^{\infty} J_{21,n}\varepsilon_n T_{2,n}^{(c)}(0)w_{1,n} - \frac{1}{l}\sum_{n=0}^{\infty} J_{22,n}\varepsilon_n T_{2,n}^{(c)}(0)w_{2,n}. \quad (57)$$

To determine the Fourier components $w_{j,n}$, multiply Eqs. (52) and (53) by $2/l\cos(s\pi x/l)$ and integrate from 0 to l, s being an integer. Thus, for the first plate

$$w_{1,s} = \frac{2}{l}\int_0^l \frac{2\hat{p}_i e^{-ik_x x}\cos(s\pi x/l)\mathrm{d}x}{S_1(k_x)} - \frac{2(G+K)}{l^2}I_{1,s}^{(f)}w_1(0) + \frac{2K}{l^2}I_{1,s}^{(f)}w_2(0) - \frac{2}{l^2}\sum_{n=0}^{\infty} J_{11,n}\varepsilon_n I_{1,s,n}^{(c)}w_{1,n} + \frac{2}{l^2}\sum_{n=0}^{\infty} J_{12,n}\varepsilon_n I_{1,s,n}^{(c)}w_{2,n} \quad (58)$$

and for the second plate

$$w_{2,s} = \frac{2K}{l^2}I_{2,s}^{(f)}w_1(0) - \frac{2K}{l^2}I_{2,s}^{(f)}w_2(0) + \frac{2}{l^2}\sum_{n=0}^{\infty} J_{21,n}\varepsilon_n I_{2,s,n}^{(c)}w_{1,n} - \frac{2}{l^2}\sum_{n=0}^{\infty} J_{22,n}\varepsilon_n I_{2,s,n}^{(c)}w_{2,n}, \quad (59)$$

where the following abbreviations have been used:

$$I_{j,s}^{(f)} \equiv \int_0^l \cos(s\pi x/l)T_j^{(f)}(x)\mathrm{d}x, \quad (60)$$

$$I_{j,s,n}^{(c)} \equiv \int_0^l \cos(s\pi x/l)T_{j,n}^{(c)}(x)\mathrm{d}x, \quad (61)$$

where $i = 1,2$. The integrals are calculated in Appendix C, as well as the integral in the first term in Eq. (58). A system of equations can now be set and solved for $w_1(0)$, $w_2(0)$ and the Fourier components $w_{j,n}$, if truncating the cosine expansion to N components. The displacement components to be solved for are arranged in a row vector ($\cdot^T$ being the transpose of the vector)

$$\mathbf{w} = [w_1(0)\ w_2(0)w_{1,0}w_{2,0}\ \ldots\ w_{1,N}w_{2,N}]^T.$$

Defining a matrix $\mathbf{A}$ and a row vector $\mathbf{P}$, see Eqs. (B1)–(B5), and solving for the displacement components $\mathbf{w}$ yields

$$\mathbf{w} = \mathbf{A}^{-1}\mathbf{P}. \quad (62)$$

Inserting these components in Eqs. (52) and (53) gives the displacement of the plates.

The solution can equivalently be written in a space harmonic form,

$$w_2 = \sum_{m=-\infty}^{\infty} W_m e^{-i\alpha_m x} \tag{63}$$

for the second plate (which is of interest in connection to sound transmission), where

$$W_m = \frac{-1}{lS_2(\alpha_m)} \left(Kw_1(0) - Kw_2(0) + \sum_{n=0}^{\infty} (J_{21,n} w_{1,n} - J_{22,n} w_{2,n}) \varepsilon_n \varsigma_n(\alpha_m) \right). \tag{64}$$

VI. EXTERIOR FLUID FIELDS AND RADIATION

We have so far found expressions for the displacement field of the plates. However, expressions for the sound transmission are also needed in order to have a prediction model.

Transmission is thus now to be determined. The transmission coefficient is defined as transmitted power divided by incident power, or in the present infinite case as the transmitted intensity divided by incident intensity,

$$\tau \equiv \frac{I_t}{I_i}.$$

The total sound intensity radiated per unit area of the plate, that is the sound intensity I_t in the direction normal to the plate, is

$$I_t = \tfrac{1}{2}\Re\{p_t v^*\} \tag{65}$$

which can be expressed as the sum of the sound intensity radiated normal to the plate by each harmonic (v being the velocity). Thus, if using the fact that $v^* = -i\omega w^*$ and that the radiated pressure can be written as a space harmonic series, the transmitted sound intensity can be calculated. Using the radiated pressure written as a space harmonic series, similar to Eqs. (63) and (64), Mace[12] derives the radiated sound intensity as

$$I_t = \frac{1}{2}\omega^3 \rho \sum_{n \in \text{radiators}} \frac{|W_n|^2}{|\gamma(\alpha_n)|}, \tag{66}$$

where

$$\gamma^2(\alpha_n) = \alpha_n^2 + k_z^2 - k^2 = (k_x + 2n\pi/l)^2 + k_z^2 - k^2$$

and where the sum is to be taken over all the radiating harmonics where $\gamma(\alpha_n) \in \mathbb{I}$, where $\mathbb{I}$ being the imaginary numbers.

The incident sound intensity normal to the plate is,

$$I_i = \tfrac{1}{2}\Re\{p_i v_i^*\}, \tag{67}$$

where v_i is the velocity in the normal direction of the incident wave,

$$v_i = \frac{-1}{i\omega\rho}\frac{\partial p_i}{\partial y} = \frac{k_y}{\omega\rho} p_i,$$

and thus for the incidence sound intensity

$$I_i = \frac{1}{2}\frac{\Re\{k_y^*\}}{\omega\rho}|p_i|^2 = \frac{1}{2}\frac{k_y}{\omega\rho}|p_i|^2, \tag{68}$$

where the last step is only correct if the incident exciting wave is a traveling wave.

From the relations (65)–(68) one may determine the transmission efficiency $\tau(\theta,\varphi)$ for each incidence angle θ and φ; the transmission efficiency is the ratio of the transmitted sound intensity I_t to the incident sound intensity I_i,

$$\tau(\theta,\varphi) = I_t(\theta,\varphi)/I_i(\theta,\varphi). \tag{69}$$

Using Eqs. (66) and (68), together with $k_y = k\cos\theta$, yields

$$\tau(\theta,\varphi) = \frac{\omega^4\rho^2}{k\cos\theta|p_i|^2} \sum_{n \in \text{radiators}} \frac{|W_n|^2}{|\gamma(\alpha_n)|}. \tag{70}$$

The statistical transmission coefficient τ_s is found as

$$\tau_s = \langle I_t(\theta,\varphi)\rangle / \langle I_i(\theta,\varphi)\rangle,$$

where $\langle\cdot\rangle$ denotes the mean. The statistical transmission coefficient therefore becomes

$$\tau_s = \frac{1}{\pi}\int_0^{2\pi}\int_0^{\pi/2} \tau(\theta,\varphi)\sin\theta\cos\theta \, d\theta \, d\varphi \tag{71}$$

and the transmission loss R dB is,

$$R = 10\log 1/\tau_s \ \text{dB}. \tag{72}$$

However, as the transmission loss is defined for all angles of incidence, the integral over all possible angles is needed, Eq. (71). This is a very time-consuming operation. The angle of incidence is therefore chosen randomly in order to approximate the integral over the incidence angles, i.e., a Monte Carlo approach. The calculation is ended when the maximum error (as compared to the last estimate of τ_s) is less than 0.01 dB 20 times in a row. The numerical calculations is made for 120 discrete frequency lines. These frequencies are spread logarithmic between 40 and 6300 Hz. A third octave band average over the frequency lines is then performed, yielding the results in Sec. VII.

VII. NUMERICAL RESULTS AND DISCUSSION

A. Data for the numerical examples

The numerical data in this section are just taken as typical estimations of material data, they are not taken from actual material measurement or adjusted to fit any result curve. Thus, the comparison with measurement in Sec. VII C is just a rough indication to the capacity of the model.

Consider a wall with 13 mm thick gypsum plates and wooden studs (45×95 mm, thus with a cavity depth $d = 95$ mm) and let the cavity not be filled with mineral wool. The studs are separated with a distance of $l = 0.6$ m. The bending stiffness of the gypsum plates is taken to be $D_1 = 520$ N m and $D_2 = 521$ N m, and the mass per unit area $m_1'' = m_2'' = 10.9$ kg/m^2. Young's modulus for the beams is $E_f = 9.8 \cdot 10^9$ Pa and the density for the beams $\rho_f = 550$ kg/m^3. The speed of sound is $c = 340$ m/s and density for air $\rho = 1.29$ kg/m^3 (the same numerical values are used in the cavity). Material damping in the beams and plates are taken to be $\eta = 0.06$ added to the Young's modulus. Material

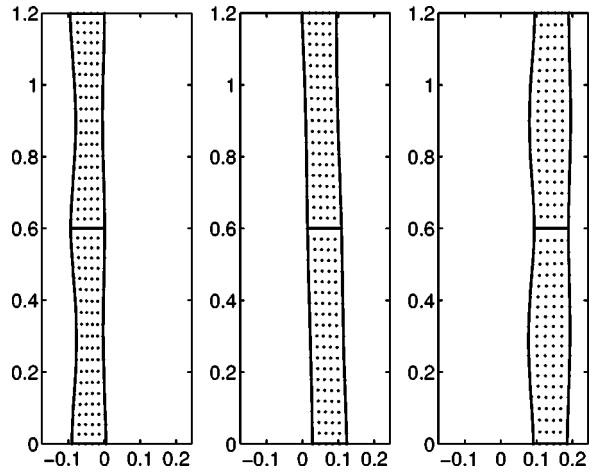


FIG. 5. Displacement of cavity field at $f=30$ Hz. Snapshots from the left: at $t=0$ s, $t=T/4$ and $t=T/2$ s, T being the period.

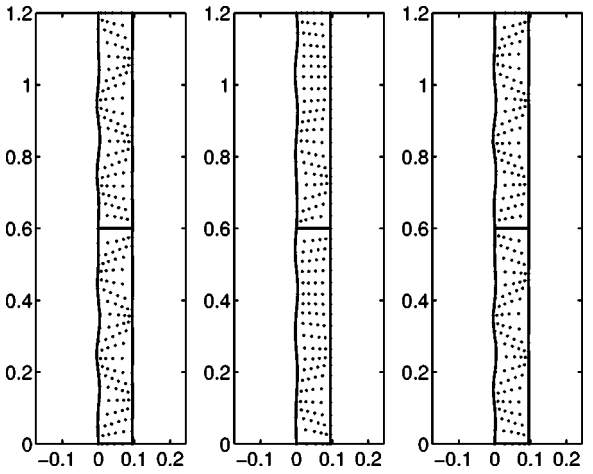


FIG. 7. Displacement of cavity field at $f=3$ kHz. Snapshots from the left: at $t=0$ s, $t=T/4$ and $t=T/2$ s, T being the period.

damping in the air is $\eta_{\text{air}}=1\cdot10^{-8}$ in the surrounding air and $\eta_{\text{air}}=1\cdot10^{-3}$ in the cavity (the reason for having higher damping in the cavity is mainly of numerical reasons, but it is also physically reasonable due to the viscous and thermal losses in the thin air layers trapped between beams and plates). The stiffness in the frame coupling is chosen to $1\cdot10^{10}\,\text{N/m}^2$ as to have a rigid connection in the frequency range of interest.

B. Displacement

The displacement of a grid of particles in the cavities is magnified and shown in Figs. 5–7. The system is excited by an incoming pressure wave with the wave numbers $k_x=k\sin(55°)$, $k_z=0$. In Fig. 5 the frequency is $f=30$ Hz. Two bays are shown and the solid lines represent the plates. Snapshots are shown, from the left representing the displacement at time $t=0$ s, $t=T/4$ and $t=T/2$ s, T being the period. In this low frequency example it can be noticed that the cavity is symmetrically deformed even though the frequency is below the first resonance. This observation appears contradictious to the result in Cremer and Heckl,[1] see pages 450–462, where a double plate system without mechanical point connectors first moves in phase which each other, and then—for frequencies above the first mass-spring-mass resonance—moves out of phase. The explanation is that the reaction fields caused by the beams are of opposite sign for the two plates, as can be seen in Eq. (41). This fact can also be seen in the next section in Cremer and Heckl,[1] see pages 462–474, where the reaction field caused by a point connector is of opposite sign in the second plate compared to the first plate. And moreover, the total displacement is in-phase as overall displacement of the wall (the rigid body motion) is larger than the more local deformation of the plates relative to the beams. In Fig. 6 the same situation is studied at $f=300$ Hz, and in Fig. 7 at $f=3$ kHz. In Fig. 6 the displacement of the two plates is truly out of phase, as the frequency is above the first resonance. The plate displacement is continuous over the stud, but the cavity field has a discontinuity, as clearly seen in the high frequency example Fig. 7. In the same figure the standing wave motion in the x-direction is also noticeable.

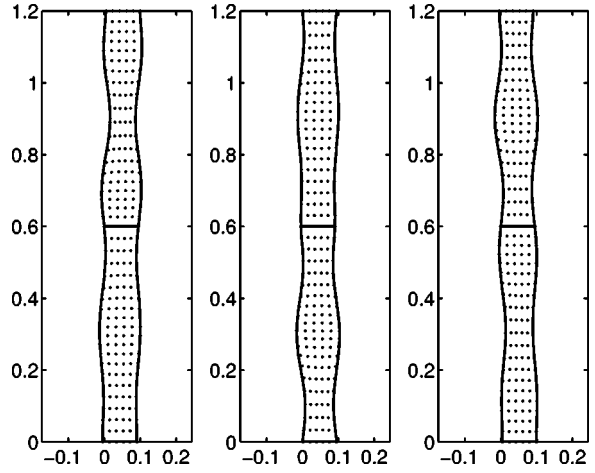


FIG. 6. Displacement of cavity field at $f=300$ Hz. Snapshots from the left: at $t=0$ s, $t=T/4$ and $t=T/2$ s, T being the period.

C. Transmission loss

The transmission loss according to Sec. VI is now to be calculated for the configuration described above. The result is shown in Fig. 8, where also experimental results are shown for the same wall (as taken from a measurement report from the 1970s found in the division archive). The agreement is satisfactory if the rough material models and the simplifications in the theoretical model are taken into account. (Note once again that the material data are estimates without special material measurements or fit to the measured transmission loss.) In the calculated transmission loss curve the troughs and peaks between 160 to 1600 Hz correspond well with those found in the experimental curve.

In Fig. 9 some variations from the base configuration are shown (relative to the base configuration, as being the one described in Sec. VII A). The effect of letting the stiffness be zero, $K=0$, is shown with circles (○). The two plates are then mechanically uncoupled. No large effects are seen, only about 2 dB are gained in the frequency interval 250–3160 Hz. It should however be noted that the airspace between the

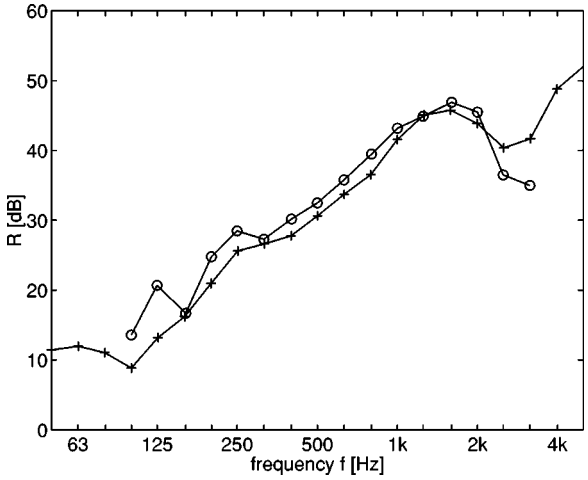


FIG. 8. Transmission loss, ○ measurements and + calculations.

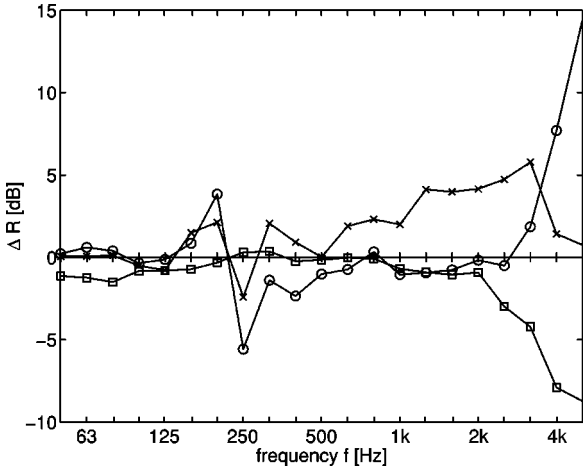


FIG. 10. Transmission loss difference relative the base configuration for the case $d=9.5$ mm, $R-R_{\text{base}}$ dB. Base configuration (+), zero stiffness $K=0$ (○), only the zero term present in cavity expansion $N=0$ (×), no bending stiffness and mass in beams (□).

plates is quite thin and undamped. The effect of only including the zero term in the expansion of the cavity is also shown (×), which is similar to having a locally reacting air cavity. Nor in this case any large effects are seen; only an increase of about 1 dB above 250 Hz and 2 dB around 2 kHz. The effect of ignoring the mass and bending stiffness of the beams is also shown (□). A small decrease at low frequencies (below 125 Hz) and a larger decrease for frequencies above 1000 Hz is seen.

In order to examine if the result in Fig. 9 is dependent on the cavity depth, more calculations were made. Thus, in Figs. 10 and 11 the same variations as in Fig. 9 are shown for cases with different cavity depth: $d=9.5$ mm in Fig. 10 and $d=950$ mm in Fig. 11. The beam height—and therefore also the beam stiffness—is the same as in the previous cases, 95 mm. The results are in general similar to Fig. 9, with some important exceptions: For the thin construction, Fig. 10, the error of using only the zero term in the cavity expansion is larger, up to 5 dB in the frequency range 1–4 kHz, compared to the other cases. For the thick construction, Fig. 11, the effect of letting the stiffness be zero, $K=0$, is about 5–8 dB for most third octave bands.

One explanation of these results can be that in the construction calculated in Fig. 9 an almost equal amount of power is transmitted via the mechanical springs as via the cavity, so if one of the transmission parts is reduced, the total transmission loss will be reduced with a maximum of 3 dB. For the really narrow cavity depth in Fig. 10 the cavity path is dominating, so here it is crucial to have an elaborate cavity description. For the thick cavity depth in Fig. 11 the mechanical path is dominating, so here the error of using only the zeroth cosine term is the smallest. The following conclusions are drawn from Figs. 9–11 concerning the number of cosine terms used: If an error of a few dB's is acceptable and if the cavity depth is not too narrow, the calculation time can be reduced without to much errors if the number of terms in the expansion is reduced. Moreover, it can also be expected that the error of letting the cavity field passing through the

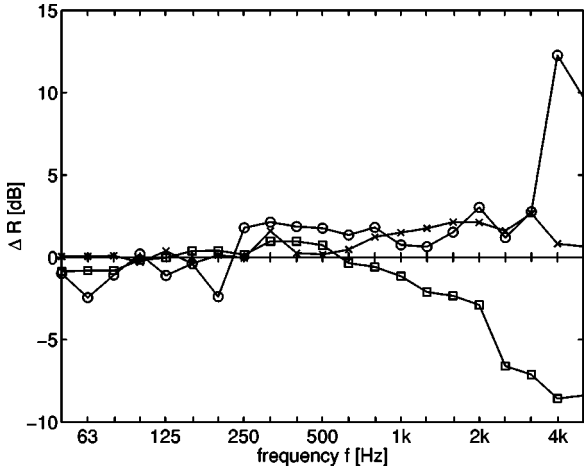


FIG. 9. Transmission loss difference relative the base configuration for the case $d=95$ mm, $R-R_{\text{base}}$ dB. Base configuration (+), zero stiffness $K=0$ (○), only the zero term present in cavity expansion $N=0$ (×), no bending stiffness and mass in beams (□).

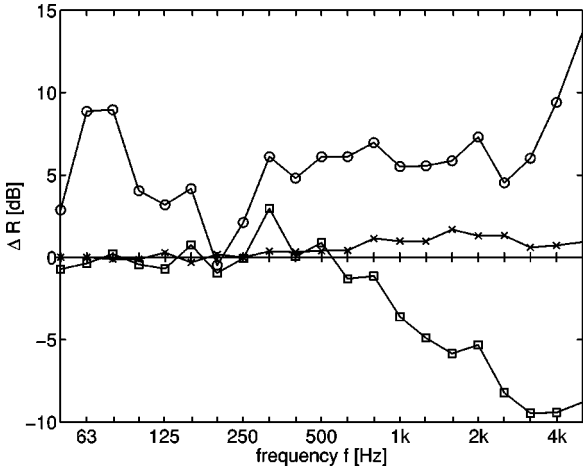


FIG. 11. Transmission loss difference relative the base configuration for the case $d=950$ mm, $R-R_{\text{base}}$ dB. Base configuration (+), zero stiffness $K=0$ (○), only the zero term present in cavity expansion $N=0$ (×), no bending stiffness and mass in beams (□).

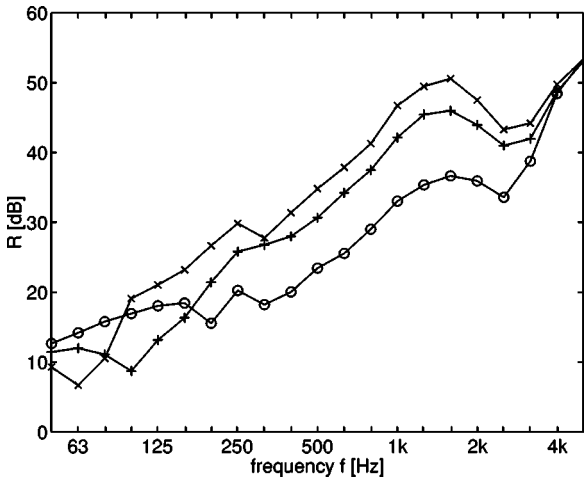


FIG. 12. Transmission loss, base configuration with different cavity depth: $d=95$ mm (+), $d=9.5$ mm (○), and $d=950$ mm (×). The beam height (and therefore also the stiffness) is the same in all three cases, 95 mm.

beams probably is of the same order of magnitude as if only the zeroth cosine term is used.

Finally, in Fig. 12 the transmission losses for the three different cavity depths $d=95$ mm, 9.5 mm, and 950 mm (but also the same data as the base configuration) is compared. The beam height is the same as in the previous cases, 95 mm. The effect on the transmission loss curve is as one can expect: increasing the cavity depths shifts the fundamental resonance frequency towards lower frequencies and increases the transmission loss above this frequency. (The term *fundamental resonance* is however a little misleading, cf. Bradley and Birta.[10])

VIII. CONCLUDING REMARKS

The paper has shown that it is possible to use a periodic assumption and transform technique to include the effects of finiteness when treating a double-plate wall with studs.

At frequencies below the first resonance, the displacements of the plates are out of phase, if the displacement relative to the beams is considered (that is if the rigid body motion is subtracted). This is due to the reaction field caused by the beams.

Comparison between measured and calculated results shows satisfactory agreement, the troughs and peaks between 160 to 1600 Hz comparing well with those being found in the experimental curve.

For the base configuration with cavity depth $d=95$ mm, no large effects are noticed if the mechanical coupling between the plates is removed or if only the zero term in the cavity expansion is used. However, for thicker constructions the mechanical coupling is important, and for thinner constructions the full expansion is important. The latter fact means that the calculation time can be reduced without to much errors if reducing the number of terms in the expansion, when the cavity is deeper than $d=95$ mm. The error of letting the cavity field passing through the beams probably is minor in such a case.

ACKNOWLEDGMENTS

The author wishes to thank his colleagues at the division of Engineering Acoustics, especially Professor Per Hammer, supervisor of this work. The KK-foundation, SBUF, and Skanska Teknik are also acknowledged for their financial support.

APPENDIX A: SIMPLIFICATION OF Φ_n

We have the following auxiliary function from Eq. (32):

$$\Phi_n=\mathcal{F}_x\left\{\cos(n\pi x/l)\sum_{m=-\infty}^{\infty}e^{-ik_x ml}(\theta(x-ml)-\theta(x-(m+1)l))\right\}.$$

Define a new function

$$\Theta_m=\mathcal{F}_x\{\cos(n\pi x/l)e^{-ik_x ml}(\theta(x-ml)-\theta(x-(m+1)l))\} \tag{A1}$$

so that

$$\Phi_n=\sum_{m=-\infty}^{\infty}\Theta_m. \tag{A2}$$

Use the following identities

$$\mathcal{F}_x\{\theta(x-a)\}=e^{i\alpha a}\left(\pi\delta(\alpha)-\frac{1}{i\varepsilon}\right), \tag{A3}$$

$$\mathcal{F}_x\{\theta(x-a)e^{ibx}\}=e^{i(\alpha-b)a}\left(\pi\delta(\alpha-b)-\frac{1}{i(\alpha-b)}\right), \tag{A4}$$

to rewrite Θ_m

$$\Theta_m=\frac{e^{-ik_x ml}}{2}\mathcal{F}_x\{\theta(x-ml)e^{in\pi x/l}-\theta(x-(m+1)l)e^{in\pi x/l}+\theta(x-ml)e^{-in\pi x/l}-\theta(x-(m+1)l)e^{-in\pi x/l}\}$$

$$=\frac{e^{-ik_x ml}}{2}\left(e^{i(\alpha-n\pi/l)ml}\left(\pi\delta(\alpha-n\pi/l)+\frac{i}{\alpha-n\pi/l}\right)-e^{i(\alpha-n\pi/l)(m+1)l}\left(\pi\delta(\alpha-n\pi/l)+\frac{i}{\alpha-n\pi/l}\right)\right)$$

$$+\frac{e^{-ik_x ml}}{2}\left(e^{i(\alpha+n\pi/l)ml}\left(\pi\delta(\alpha+n\pi/l)+\frac{i}{\alpha+n\pi/l}\right)-e^{i(\alpha+n\pi/l)(m+1)l}\left(\pi\delta(\alpha+n\pi/l)+\frac{i}{\alpha+n\pi/l}\right)\right).$$

Taking the limit $\alpha \to \pm n\pi/l$ it can be shown that the amplitude of Dirac's function equals zero. Thus, after some rearrangement

$$\Theta_m = -\frac{e^{-ik_x ml} e^{i\alpha ml}}{2}\left((1-e^{i\alpha l}e^{-in\pi}) \frac{e^{-inm\pi}}{i(\alpha - n\pi/l)} + (1-e^{i\alpha l}e^{in\pi}) \frac{e^{inm\pi}}{i(\alpha + n\pi/l)} \right). \tag{A5}$$

Therefore, using Eqs. (A2) and (A5),

$$\Phi_n = -\frac{1}{2i}\left(\frac{1-e^{i\alpha l}e^{-in\pi}}{\alpha - n\pi/l} \sum_{m=-\infty}^{\infty} e^{-inm\pi} e^{i(\alpha - k_x)ml} + \frac{1-e^{i\alpha l}e^{in\pi}}{\alpha + n\pi/l} \sum_{m=-\infty}^{\infty} e^{inm\pi} e^{i(\alpha - k_x)ml} \right). \tag{A6}$$

The Poisson sum formula can be used to show that

$$\sum_{m=-\infty}^{\infty} f(am) = \frac{1}{a} \sum_{q=-\infty}^{\infty} F\left(\frac{2q\pi}{a} \right), \tag{A7}$$

where F is the Fourier transform (8) of the arbitrary function f, which together with the rules for the Dirac function $\delta(\alpha l) = \delta(\alpha)/l$ implies that

$$\sum_{m=-\infty}^{\infty} e^{i(\alpha - k_x)ml} e^{\pm in\pi m} = \frac{2\pi}{l} \sum_{m=-\infty}^{\infty} \delta(\alpha - k_x + (n \pm 2m)\pi/l). \tag{A8}$$

APPENDIX B: THE MATRIX COMPONENTS

The matrix components in (62) are

$$A_{11} = 1 + \frac{(G+K)}{l} T_1^{(f)}(0),$$

$$A_{12} = -\frac{K}{l} T_1^{(f)}(0), \tag{B1}$$

$$A_{1n} = \begin{cases} \dfrac{J_{11,n}\varepsilon_n}{l} T_{1,n}^{(c)}(0), & n \text{ odd} \\ \dfrac{-J_{12,n}\varepsilon_n}{l} T_{1,n}^{(c)}(0), & n \text{ even} \end{cases}$$

and

$$A_{21} = -\frac{K}{l} T_2^{(f)}(0),$$

$$A_{22} = 1 + \frac{K}{l} T_2^{(f)}(0) \tag{B2}$$

$$A_{2n} = \begin{cases} -\dfrac{J_{21,n}\varepsilon_n}{l} T_{2,n}^{(c)}(0), & n \text{ odd} \\ \dfrac{J_{22,n}\varepsilon_n}{l} T_{2,n}^{(c)}(0), & n \text{ even} \end{cases}$$

and for odd s

$$A_{s1} = \frac{2(G+K)}{l^2} I_{1,s}^{(f)},$$

$$A_{s2} = -\frac{2K}{l^2} I_{1,s}^{(f)}, \tag{B3}$$

$$A_{sn} = \delta_{sn} + \begin{cases} \dfrac{2}{l^2} J_{11,n}\varepsilon_n I_{1,s,n}^{(c)}, & n \text{ odd} \\ -\dfrac{2}{l^2} J_{12,n}\varepsilon_n I_{1,s,n}^{(c)}, & n \text{ even,} \end{cases}$$

where δ_{sn} is the Kronecker delta that equals one if $s=n$ and equals zero otherwise, and for even s

$$A_{s1} = -\frac{2K}{l^2} I_{2,s}^{(f)},$$

$$A_{s2} = \frac{2K}{l^2} I_{2,s}^{(f)}, \tag{B4}$$

$$A_{sn} = \delta_{sn} + \begin{cases} -\dfrac{2}{l^2} J_{21,n}\varepsilon_n I_{2,s,n}^{(c)}, & n \text{ odd} \\ \dfrac{2}{l^2} J_{22,n}\varepsilon_n I_{2,s,n}^{(c)}, & n \text{ even.} \end{cases}$$

The vector components in (62) are

$$P_1 = \frac{2\hat{p}_i}{S_1(k_x)},$$

$$P_2 = 0, \tag{B5}$$

$$P_s = \begin{cases} \dfrac{2 \cdot 2\hat{p}_i}{lS_1(k_x)} \displaystyle\int_0^l e^{-ik_x x} \cos(s\pi x/l)\,dx, & s \text{ odd} \\ 0, & s \text{ even,} \end{cases}$$

where the last expression is valid for $s>2$, and where the integral is

$$\int_0^l e^{-ik_x x} \cos(s\pi x/l)\,dx = ik_x \frac{1-(-1)^s e^{-ik_x l}}{(s\pi/l)^2 - k_x^2}.$$

APPENDIX C: CALCULATION OF INTEGRALS

The integral in Eq. (60) is, using Eq. (54) and interchanging the integral and the sum

$$I_{i,s}^{(f)} = \sum_{m=-\infty}^{\infty} \frac{I_s(\alpha_m)}{S_i(\alpha_m)}, \tag{C1}$$

where

$$I_s(\alpha_m) = \int_0^l \cos(s\pi x/l) e^{-i\alpha_m x}\,dx = i\alpha_m \frac{1-(-1)^s e^{-i\alpha_m l}}{(s\pi/l)^2 - \alpha_m^2}. \tag{C2}$$

The integral in Eq. (61) is, using Eq. (55) and interchanging the integral and the sum

$$I_{i,s,n}^{(c)} = \sum_{m=-\infty}^{\infty} \varsigma_n(\alpha_m) \frac{I_s(\alpha_m)}{S_i(\alpha_m)}, \tag{C3}$$

where $I_s(\alpha_m)$ is evaluated according to Eq. (C2).

[1] L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound* (Springer-Verlag, Berlin, 1988).
[2] EN 12354: Building acoustics—estimation of acoustic performance of buildings from the performance of products. Part I: Airborne sound insulation between rooms. EN 12354-1. Part II: Impact sound insulation between rooms. EN 12354-2, European Standard, 1996.
[3] T. R. T. Nightingale and I. Bosmans, ''Vibration response of lightweight wood frame building elements,'' J. Building Acoustics **6**, 269–288 (1999).
[4] R. J. M. Craik, *Sound Transmission Through Buildings* (Gower, Aldershot, Hampshire GU11 3HR, England, 1996).
[5] R. J. M. Craik and R. S. Smith, ''Sound transmission through double leaf lightweight partitions part I: Airborn sound,'' Appl. Acoust. **61**, 223–245 (2000).
[6] B. H. Sharp, ''Prediction methods for the sound transmission of building elements,'' Noise Control Eng. **11**, 53–63 (1978).
[7] A. London, ''Transmission of reverberant sound through double walls,'' J. Acoust. Soc. Am. **22**, 270–279 (1950).
[8] W. Kropp and E. Rebillard, ''On the air-borne sound insulation of double wall constructions,'' Acust. Acta Acust. **85**, 707–720 (1999).
[9] G. F. Lin and J. M. Garrelick, ''Sound transmission through periodically framed parallel plates,'' J. Acoust. Soc. Am. **61**, 1014–1018 (1977).
[10] J. S. Bradley and J. A. Birta, ''On the sound insulation of wood stud exterior walls,'' J. Acoust. Soc. Am. **110**, 3086–3096 (2001).
[11] M. L. Rumerman, ''Vibration and wave propagation in ribbed plates,'' J. Acoust. Soc. Am. **57**, 370–373 (1975).
[12] B. R. Mace, ''Periodically stiffened fluid-loaded plates, I: Response to convected harmonic pressure and free wave propagation,'' J. Sound Vib. **73**, 473–486 (1980).
[13] D. Takahashi, ''Sound radiated from periodically connected double-plate structures,'' J. Sound Vib. **90**, 541–557 (1983).
[14] E. A. Skelton, ''Acoustic scattering by parallel plates with periodic connectors,'' Proc. R. Soc. London, Ser. A **427**, 419–444 (1990).
[15] K. Sakagami, H. Gen, M. Morimoto, and D. Takahashi, ''Acoustic reflection of an elastic plate with periodic ribs and a back cavity,'' Acust. Acta Acust. **85**, 1–11 (1999).
[16] E. A. Skelton, ''Acoustic scattering by a rigid barrier between two fluid-loaded parallel plates,'' Proc. R. Soc. London, Ser. A **435**, 217–232 (1991).
[17] I. A. Urusovskii, ''Sound transmission through two periodically framed parallel plates,'' Sov. Phys. Acoust. **38**, 411–413 (1992).
[18] J. Brunskog and P. Hammer, ''Prediction model for the impact sound level of lightweight floors,'' Acust. Acta Acust. **89**, 309–322 (2003).
[19] W. Maysenhoelder and R. Haberkern, ''Sound transmission through periodically inhomogeneous plates: Solution of the general problem by a variational formulation,'' Acust. Acta Acust. **89**, 323–332 (2003).
[20] J. Brunskog and P. Hammer, ''Prediction models of impact sound insulation on timber floor structures; A literature survey,'' J. Building Acoustics **7**, 89–112 (2000).

A novel method for estimating the focal size of two confocal high-intensity focused ultrasound transducers

Wen-Shiang Chen
Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei 100, Taiwan and Department of Mechanical Engineering, National Taiwan University, Taipei 106, Taiwan

Ping-Mo Ma
Department of Mechanical Engineering, National Taiwan University, Taipei 106, Taiwan

Hao-Li Liu
Department of Electronic Engineering, Tung-Nan Institute of Technology, Taipei, Taiwan

Chih-Kuang Yeh[a)]
Department of Electrical Engineering, Yuan Ze University, Taoyuan 320, Taiwan

Min-Shin Chen
Department of Mechanical Engineering, National Taiwan University, Taipei 106, Taiwan

Chein-Wei Chang
Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei 100, Taiwan

(Received 7 November 2004; revised 4 March 2005; accepted 15 March 2005)

Estimating the focal size and position of a high-intensity focused ultrasound (HIFU) transducer remains a challenge since traditional methods, such as hydrophone scanning or schlieren imaging, cannot tolerate high pressures, are directional, or provide low resolution. The difficulties increase when dealing with the complex beam pattern of a multielement HIFU transducer array, e.g., two transducers facing each other. In the present study we show a novel approach to the visualization of the HIFU focus by using shockwave-generated bubbles and a diagnostic B-mode scanner. Bubbles were generated and pushed by shock waves toward the HIFU beam, and were trapped in its pressure valleys. These trapped bubbles moved along the pressure valleys and thereby delineated the shape and size of the HIFU beam. The main and sidelobes of 1.1- and 3.5 MHz HIFU beams were clearly visible, and could be measured with a millimeter resolution. The combined foci could also be visualized by observing the generation of sustained inertial cavitation and enhanced scattering. The results of this study further demonstrate the possibility of reducing the inertial cavitation threshold by the local introduction of shock wave-generated bubbles, which might be useful when bubble generation and cavitation-related bioeffects are intended within a small region *in vivo*. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1904283]

PACS numbers: 43.58.Vb, 43.25.Cb, 43.35.Ei [RR] Pages: 3740–3749

I. INTRODUCTION

The minimally invasive nature of high-intensity focused ultrasound (HIFU) has been a focus of several recent therapeutic applications, especially tumor ablation.[1–5] The temperature of the target tissue at the focus of a HIFU transducer can increase to more than 65 °C within seconds, thus causing protein denaturing and cell death. However, the success of HIFU tumor ablation largely depends on the knowledge of the focusing properties of the treatment system, making techniques for quantitatively measuring the focal position and size essential.

Typical tools for measuring the position and size of a HIFU transducer are hydrophone scanning and schlieren optical imaging.[6] A computer-controlled scanning system with a needle or membrane hydrophone can accurately determine the position, size, and pressure amplitude of the main and sidelobes of a HIFU sound field. Although a very fine resolution can be achieved, the scanning process is time consuming and the alignment of the hydrophone to the HIFU transducer is critical to the accuracy of the measurements. Moreover, the hydrophone is fragile and quickly saturated at high pressures, and the calibration of a hydrophone in a multielement HIFU system becomes difficult due to the acoustic waves coming from several sources differing greatly in their incoming angles. Schlieren optic imaging provides an overview of the shape and location of a HIFU focus as well as sidelobes and secondary foci. However, the resolution is low and thus the obtained results are more qualitative than quantitative.

In this article we report a novel method for estimating the position and size of the focus of a 1.1 or 3.5 MHz HIFU transducer. The shapes and sizes of the main and sidelobes can be determined in a reasonably short time and with a high resolution. Furthermore, the changes in shape and pressures

[a)]For correspondence: Dr. Chih-Kuang Yeh, Department of Electrical Engineering, Yuan Ze University, 135 Yuan-Tung Rd., Chung-Li, 320 Taoyuan, Taiwan. Telephone: +886-3-463-8800, ext. 2438 (voice); +886-3-463-9355 (fax); electronic mail: ckyeh@saturn.yzu.edu.tw

 0001-4966/2005/117(6)/3740/10/$22.50

TABLE I. Simulation parameters.

Frequency (f)	1.1 MHz	3.5 MHz
Transducer radius	32 mm	17.5 mm
Geometric focus (data from the manufacturer)	62.64 mm	55 mm
Attenuation	0 Np/cm	0 Np/cm
Sound speed (c)	1500 m/s	1500 m/s
Wavelength (λ)	1.364 mm	0.429 mm
Grid size		
Field	0.1 mm	0.1 mm
Transducer (Δx and Δy)	0.164 mm	0.164 mm

where the foci of the two transducers intersect are also determined.

II. EXPERIMENTAL MATERIALS AND METHODS

A. Experimental setup

The experiments were performed in an acrylic tank (23 cm L×15 cm W×15 cm H) filled with tap water. The experimental system consisted of four transducers (1.1 and 3.5 MHz HIFU transducers; H-101 and SU-102, Sonic Concepts, Woodinville, USA), a shock wave transducer (Piezoson 100 with a FB7 G2 probe, Richard Wolf, Knittlingen, Germany) for producing bubbles, and a diagnostic B-mode scanner (Titan with a L38 probe, Sonosite, Bothell, USA) for imaging. The L38 is a 5–10 MHz broadband transducer with a central frequency of 7.5 MHz. The focus of the probe used with the shock wave generator was 1.7 mm in diameter and 6.7 mm in length, while the focal length was 20 mm from the edge plane of the probe (data from the manufacturer). The diameters and focal lengths of the 1.1 and 3.5 MHz transducers are listed in Table I. The 1.1 MHz HIFU and the shock wave transducers were positioned orthogonally on two adjacent walls of the water tank, as shown in Fig. 1. For experiments on the combined beam pattern of two HIFU transducers, the second HIFU transducer (3.5 MHz) was mounted on the wall

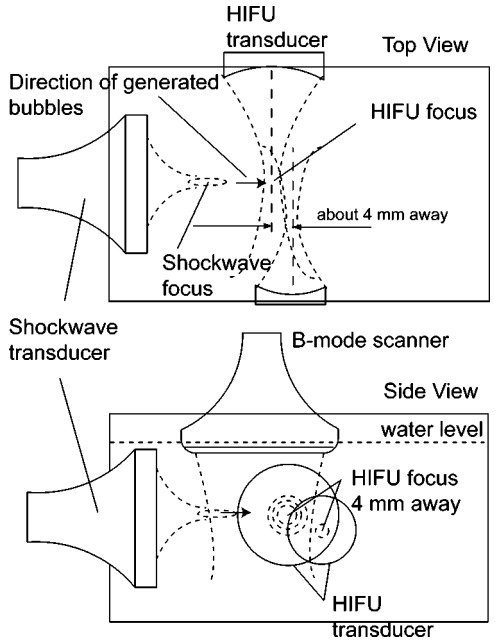


FIG. 1. Setup of the experimental apparatus: top and side views.

facing the 1.1 MHz HIFU transducer with their central axes parallel to each other but separated by about 4 mm. The B-mode probe scanned from the top of the water tank parallel to the axial direction of the shock wave transducer, i.e., cutting through the HIFU beam. The HIFU transducers were both mounted on 3-D positioners that allowed precise movement control.

The pressure values of the HIFU transducers were calculated from hydrophone (Onda, Sunnyvale, USA) pressure measurements performed in water at low output amplitudes that were linearly extrapolated to higher outputs. The peak negative pressures at the focus of the 1.1 MHz transducer were 1.62 and 5.88 MPa for outputs of 100 and 400 mV from the function generator, respectively. The peak negative pressure of the 3.5 MHz transducer was around 3 MPa at the focal point. Two arbitrary function generators (33120A, Agilent, Palo Alto, USA; and DS345, Stanford Research Systems, Sunnyvale, USA) and two power amplifiers (150A250B and 150A100B, Amplifier Research, Souderton, USA) were used to drive the 3.5 and 1.1 MHz HIFU transducers, respectively. The pulse repetition frequency (PRF) was 500 Hz, and the pulse length was 100 cycles for both transducers. A clinical shock wave generator was used to induce inertial cavitation and generate bubbles, and was set to its highest output level (20, the corresponding peak negative pressure provided by the manufacturer was 19.0 MPa) and the highest PRF (4 Hz) for maximal bubble production. The beam-pattern images obtained by the diagnostic B-mode scanner were recorded on a VCR and off-line processed frame-by-frame using commercial software (Premiere Pro 1.5, Adobe Systems, San Jose, USA; Matlab, The MathWorks, Natick, USA).

B. Experimental procedures

1. Series 1. Beam pattern of a single HIFU transducer

In order to test the feasibility of using shock wave-generated bubbles to visualize the beam pattern of HIFU transducers, the 1.1 and 3.5 MHz transducers were first tested individually. The 1.1 MHz HIFU transducer was driven at either 100 or 400 mV and moved along its axial direction at a step size of 1 or 2 mm, respectively. The shock wave generator produced and pushed bubbles toward the HIFU beam while B-mode images were taken continuously. After finishing ''slicing'' along the HIFU beam, the B-mode imaging plane was aligned with the HIFU focal plane again, and B-mode and color Doppler images were both taken. The output of the 1.1 MHz HIFU transducer was increased until enhanced scattering due to inertial cavitation at the center of the HIFU beam was induced. The beam pattern of the 3.5 MHz transducer was determined in a similar way.

2. Series 2. Simulation

To understand the beam profiles obtained in the above experiments, simulations were performed by modeling the 1.1 MHz HIFU transducer as a collection of point sources in a grid. For any point on the destination plane (fixed z value),

the following Rayleigh–Sommerfold integral[7] was used to calculate the relative pressure amplitude contributed by each grid point of the source transducer:

$$p(x,y,z) = i\frac{\rho c k}{2\pi}\int_S \frac{u e^{-ik(r-r')}}{r-r'} dS, \tag{1}$$

where $i=\sqrt{-1}$, ρ=tissue density, c=sound speed, k=wave number ($k=2\pi/\lambda$, λ is the wavelength, and $c=f\lambda$, where f is sound frequency), u=complex surface velocity of source, and $r-r'$=distance between a certain point on the transducer surface to a certain point in the acoustic field. For more efficient and convenient calculations, the above equation was simplified to

$$p(x,y,z) = i\,\mathrm{AMP}\,\frac{\Delta x\,\Delta y}{2\pi}\sum\sum\frac{e^{-ik(r-r')}}{r-r'}. \tag{2}$$

The amplitude of the simulated pressure field is relative when AMP is set to 1, and Δx and Δy are the step sizes in the x and y directions. The parameters used in the simulation program are listed in Table I. The simulated results were then compared with the experimental results (see Sec. III).

3. Series 3. Combined beam pattern of two HIFU transducers

The major challenge in beam-pattern plotting is encountered during the testing of a multielement transducer (especially when the elements are at large angles to each other). To demonstrate the advantages of our newly developed method, two HIFU transducers (one 1.1 and the other 3.5 MHz) were arranged facing each other, but with a 4 mm distance between their parallel central axes on purpose. The 3.5 MHz transducer was turned on first since its focal pattern was small and difficult to determine. The obtained B-mode images of the overlapping beam pattern were recorded and analyzed offline.

III. RESULTS

A. Beam pattern of a single HIFU transducer

To rapidly determine the focal size and location of a HIFU beam, the HIFU transducer was moved along its axial direction while B-mode images were taken sequentially. Figure 2(a) shows one of the recorded images near the focus of a 1.1 MHz HIFU beam. When bubbles produced by the shock wave transducer were pushed toward the HIFU beam, a few concentric rings (white rings) appeared where bubbles were trapped and moved along the rings, which clearly visualized the locations of the pressure valleys around the HIFU focus (Fig. 3). The diameter of the central dark region was smaller for the 3.5 MHz transducer than for the 1.1 MHz transducer (Fig. 4). The radius of the first ring, which can be easily measured, was the radius of the first zero of the acoustic beam, and thus was larger than the focal radius (usually defined as being 6 dB below the maximum pressure) of the HIFU beam. In other words, the focal radius should be smaller than the diameter of measured radius of the first ring, 1.6 mm in Fig. 5(a). The number of visible rings increased with the pressure level [Fig. 2(b) and Fig. 5], whereas the diameter of the first ring was the same for low [Fig. 5(a)] and high [Fig. 5(b)] pressures. The minimum diameter of the central dark region occurred 48 mm from the edge plane of the 1.1 MHz HIFU transducer [Figs. 5(a) and 5(b)]. At a low-pressure condition, ''radius steps'' were seen at a few positions, e.g., 40–41, 42–45, 46–50, 50–54, and 55–59 mm [Fig. 5(a)]. At the transition from one step to the next one, the measured radii could be either at the upper ''step'' (case 2) or the lower ''step'' (case 1). The radius of the second ring (white triangle) for case 1 coincided with that of the first ring for case 2 (gray square).

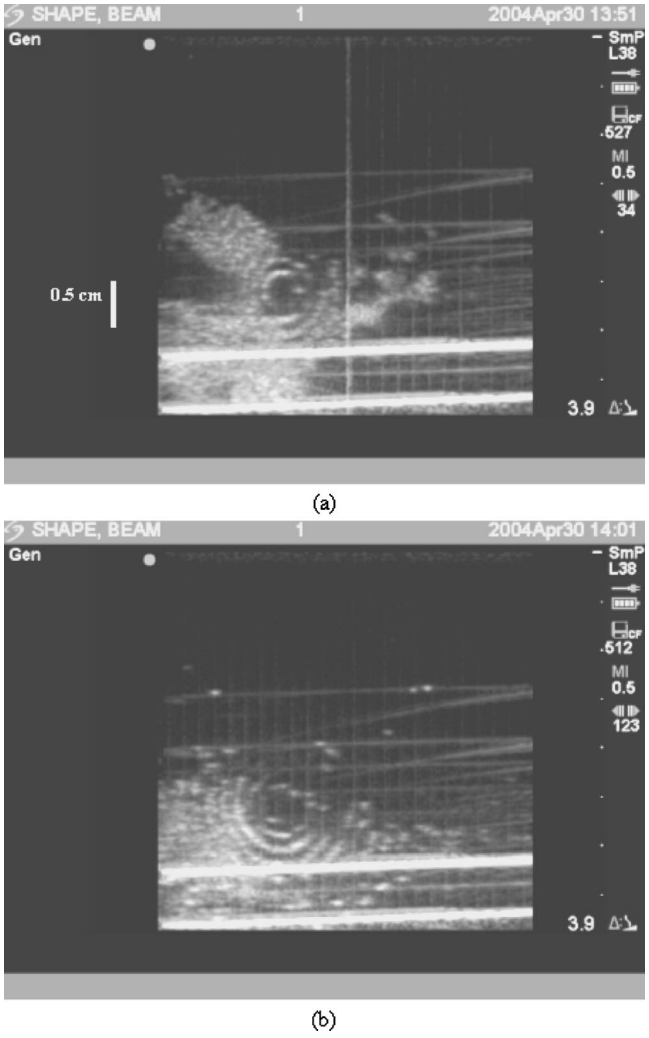


FIG. 2. (a) A B-mode image of the beam pattern of a 1.1 MHz HIFU transducer near its focus. The output from the function generator is 100 mV (producing a peak negative pressure of 1.62 MPa). The vertical white line was produced by a shock wave pulse. (b) A B-mode image of the beam pattern of a 1.1 MHz HIFU transducer at the same position as (a), but for a higher voltage output from the function generator (400 mV, 5.88 MPa). More rings are clearly evident.

Bubbles for visualizing the beam pattern of a HIFU transducer can be provided by means other than a shock wave transducer. Figure 6 shows the bubbles generated at the center of the HIFU focus by inertial cavitation at higher pressure levels (frame 3) that visualize the first and second rings of the HIFU beam when they moved outward. However, compared to the bubbles generated by a shock wave trans-

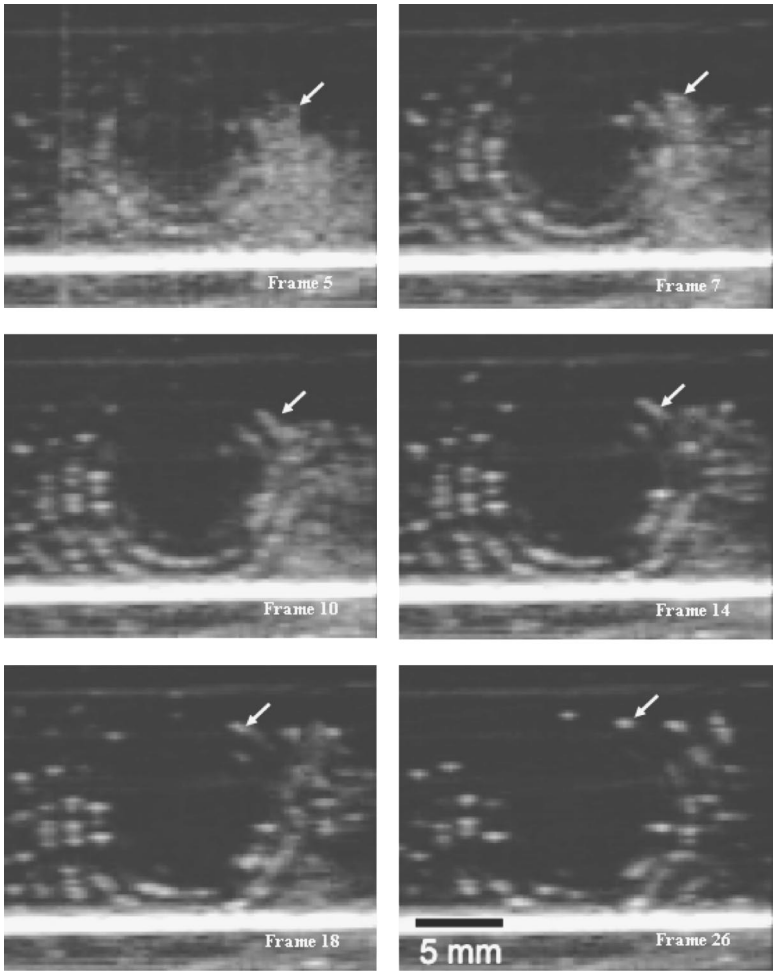


FIG. 3. A group of bubbles (white solid arrow) is moving along the third pressure valley (the third ring) of the 1.1 MHz HIFU beam. The acoustic parameters used in this test was 100 mV output from the function generator, pulse length=100 cycles, and PRF=500 Hz. The B-mode imaging slice cut through the focal plane of the HIFU transducer (48 mm from the transducer's edge plane). In this figure, only frames 5, 7, 10, 14, 18, and 26 are shown for simplicity.

ducer, bubbles from the HIFU focus lasted for a shorter time and thus were probably smaller. Figure 7 shows a Doppler image of the bubbles generated by a HIFU transducer. An oscillating ball with rapidly changing color was evident at the focus of the HIFU transducer, which is indicative of the fast phase change in backscattered signals due to the generation and collapse of bubbles by inertial cavitation. This phenomenon was sometimes called the pseudo-Doppler shift and has been used to produce images of "stimulated acoustic emission."[8]

B. Matching experimental and simulated results

In order to elucidate the physical meaning of the experimental results, the beam patterns of both 1.1 and 3.5 MHz HIFU transducers were simulated and compared to the rings evident on the B-mode images. In Fig. 8, it is clear that the white rings of the 1.1 MHz HIFU beam matched well with the location of the pressure valleys measured using a needle hydrophone. It is worth noting that the pressures at the valleys were not zero. The measured pressure might be the sum

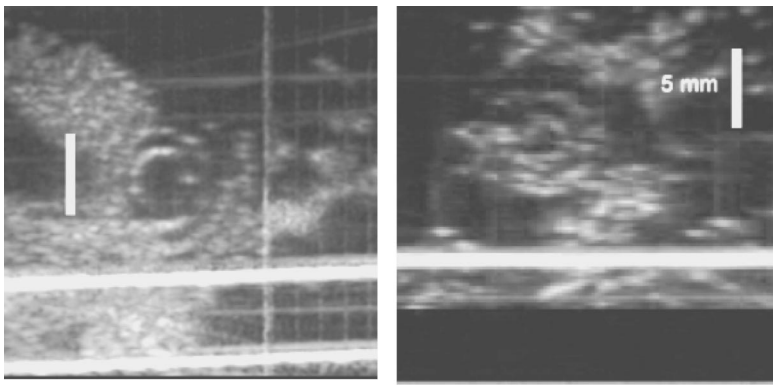


FIG. 4. The focal beam patterns of (a) 1.1 MHz and (b) 3.5 MHz HIFU transducers. The B-mode imaging slice cut through the focal planes of both HIFU transducers. Comparing (a) and (b), the focal beam size of the 3.5 MHz transducer is smaller than that of the 1.1 MHz one.

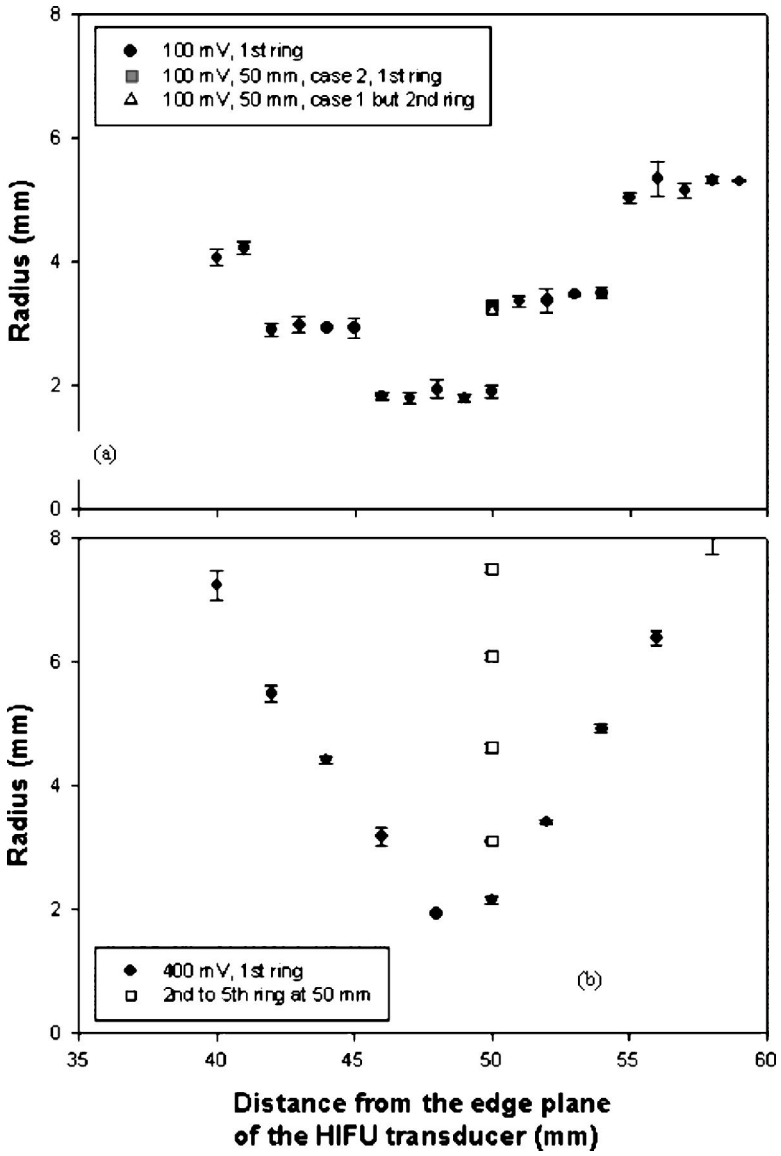


FIG. 5. The radii of each ring at (a) low- (100 mV) and (b) high-pressure (400 mV) output are shown. The minimum diameter of the 1.1 MHz HIFU transducer occurred 48 mm from the edge plane. At a low-pressure condition, ''radius steps'' were seen at a few positions, e.g., from 46 to 50 mm and from 50 to 54 mm. At the transition between two steps, e.g., at 50 mm, the measured radii could be either at the upper ''step'' (case 2, gray square) or the lower ''step'' (case 1, black circle). The radius of the second ring (white triangle) for case 1 coincided with that of the first ring for case 2 (gray square). The error bar indicates the size of the standard deviation. At least 3 measurements were performed for each data point.

of transmitted and reflected waves from all directions. Furthermore, for the 1.1 MHz HIFU transducer, the locations of the dark rings in the images match well the locations of the simulated pressure peaks [Fig. 9(a)]. Similarly, the locations of the first white ring coincide well with the location of the second pressure valley for the 3.5 MHz HIFU transducer. The tiny inner ring in the simulation results cannot be seen in the experiment, probably because of the resolution limitations of B-mode images [Fig. 9(b)].

C. The combined beam pattern of two HIFU transducers

To understand the beam pattern of two facing transducers, the focal location of each transducer was determined first. The focus of the 1.1 MHz transducer was then moved by adjusting the 3-D positioner to make the focal planes of both the 1.1 and 3.5 MHz beams coincide. The two foci were separated by 4 mm distance on purpose.

The pressure levels of the 1.1 and 3.5 MHz HIFU transducers were both below the pressure threshold to induce inertial cavitation of bubbles provided by shock waves. Therefore, before turning both transducers on, no bubble was generated near the confocal plane of both transducers after the bubbles from the shock waves passed or dissolved. When two transducers were turned on simultaneously, the pressures of both transducers summed up and thus the peak pressure level near the focus of the 3.5 MHz transducer increased. The bubbles from the shock wave transducer probably acted as seeds of inertial cavitation and induced bubble generation. Continuous bubble generation was seen at the intersecting area and could persist up to a few minutes when bubbles from the shock wave generator disappeared for a long time. The enhanced scattering signals of the generated bubbles then clearly demonstrated the location of the intersecting area (Fig. 10).

IV. DISCUSSION

In this report we describe the results of a simple and novel method to determine the size and location of HIFU beams. The new method is particularly useful when analyzing a transducer array with a complex arrangement, for example, as used in the noninvasive ablation of a brain tumor

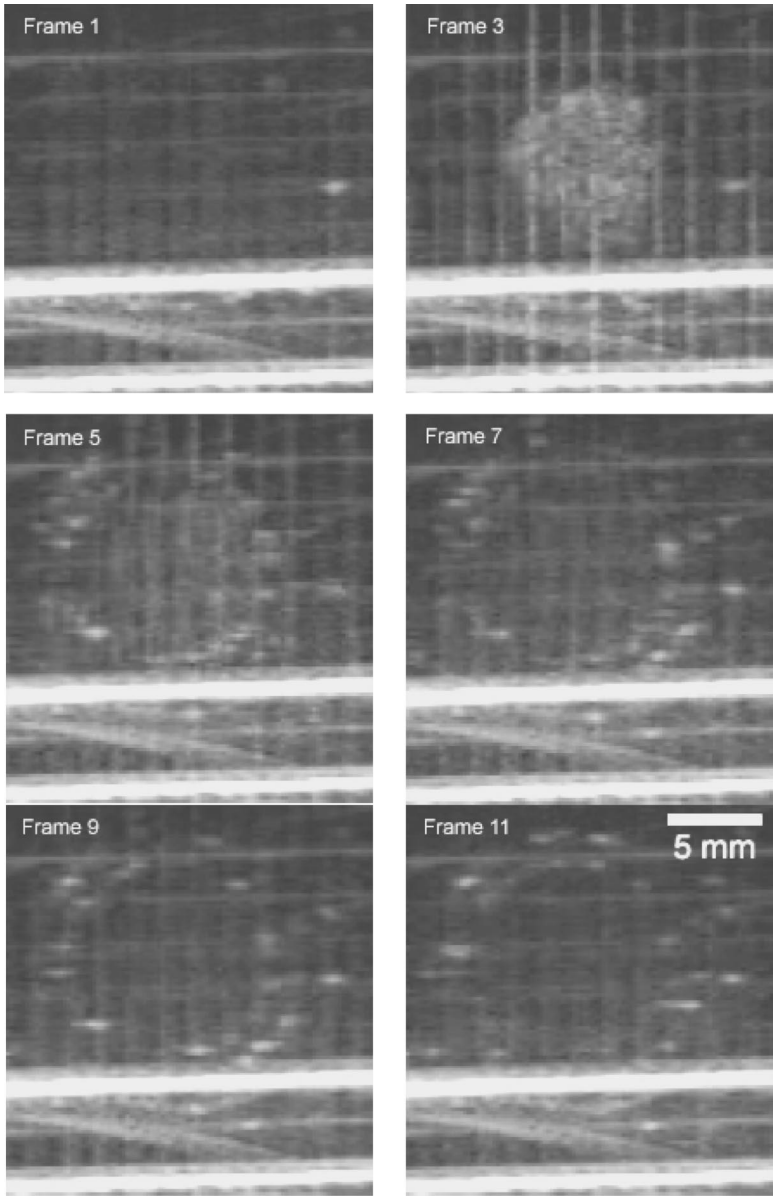


FIG. 6. The beam pattern of a 1.1 MHz HIFU transducer was visualized by bubbles generated by inertial cavitation (frame 3). Two rings are clearly evident in frames 5–11 (550 mV or P=8.09 MPa, 100 cycles, 500 Hz PRF). The vertical white lines in each frame were produced by the HIFU transducer itself. Only frames 1, 3, 5, 7, 9, and 11 are shown.

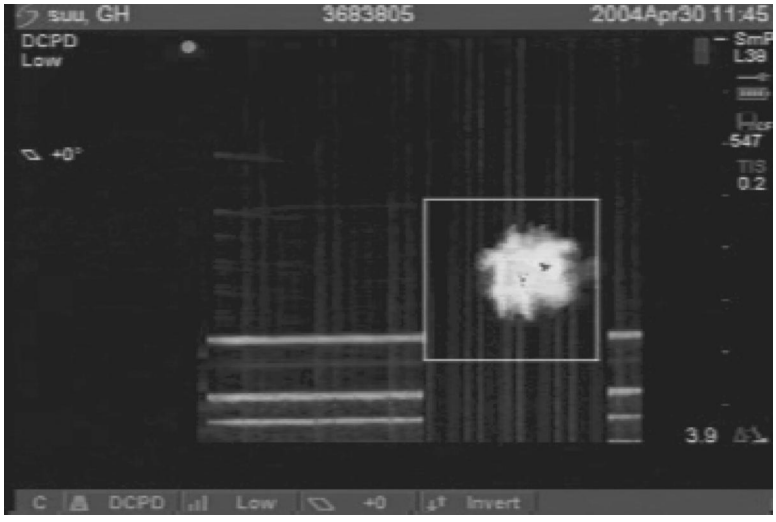


FIG. 7. The bubbles generated by inertial cavitation at the focus of a 1.1 MHz HIFU transducer looks like a ball changing in color rapidly (mosaic pattern) in a directional color power Doppler image (550 mV, 100 cycles, 500 Hz PRF).

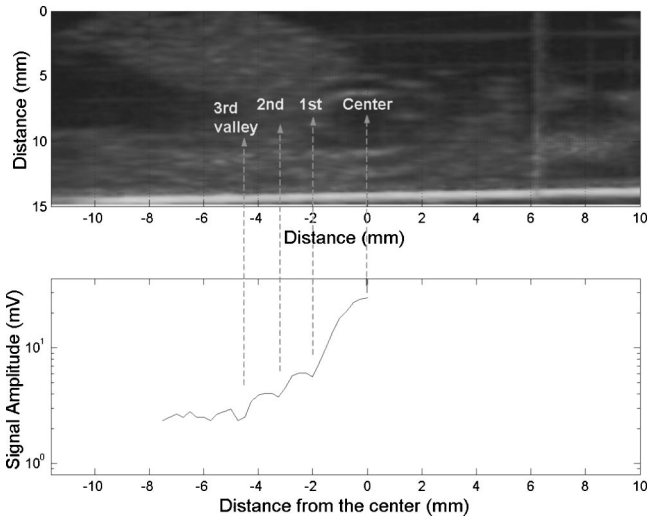


FIG. 8. The white rings of the 1.1 MHz HIFU beam compared with the location of the pressure valleys measured by a needle hydrophone.

using a transcranial ultrasound array. Arranging two transducers from the bilateral temporal area of the skull—where the bone is thinner—may be preferable because the output from each transducer decreased and the possibility of overheating the brain tissue outside the target reduced. Unfortunately, the resulting beam pattern cannot be measured by a traditional needle hydrophone system. Furthermore, our proposed method can also be used to rapidly construct a 3-D beam pattern, in contrast to 3-D measurements with a needle hydrophone system usually taking many hours to complete. The proposed method is able to perform beam plotting of two facing transducers, and can construct a 3-D beam pattern in a relatively shorter time. However, the proposed method is not able to determine the absolute peak pressure or intensity, or the pressure profile of the focus. In addition, the resolution is limited by the frequency of the diagnostic ultrasound probe (in our case, a 7.5 MHz central frequency), and thus is not adequate for a focal size smaller than 1 mm. To clearly show a whole ring (the first pressure valley) in the B-mode images, an area of about 1×1 mm^2 (or about 5×5 pixels) is necessary. The use of a diagnostic probe with a higher frequency would increase the resolution.

Before turning on both transducers, the output of each transducer was lower than the inertial cavitation threshold of the bubbles from the shock wave generator, and hence no enhanced scattering was seen. When the foci of two transducers intersected, the peak pressure increased and exceeded these bubbles' inertial cavitation threshold. Therefore, before introducing bubbles to the intersecting area, no inertial cavitation or enhanced scattering was detected. Seeding some

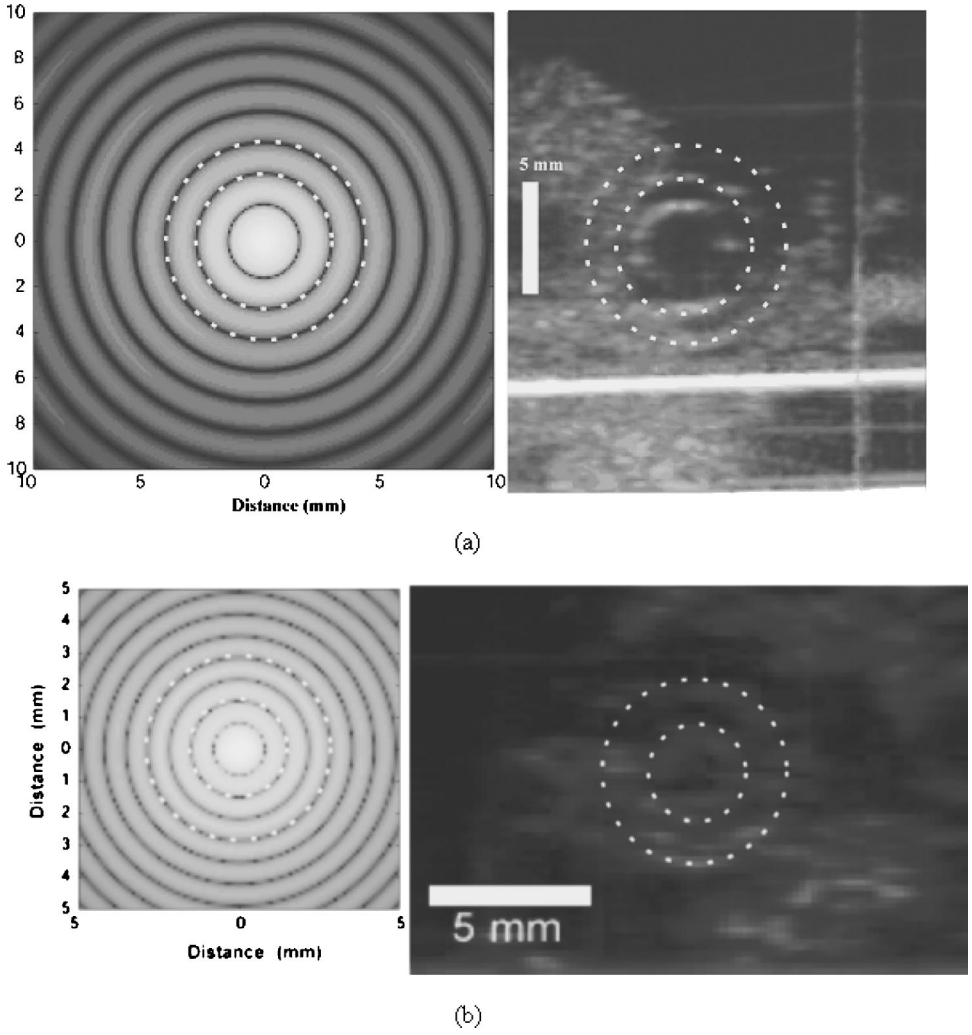


FIG. 9. The comparisons between the simulation and experimental results are shown for (a) the 1.1 MHz and (b) the 3.5 MHz HIFU transducers. For both transducers, the B-mode imaging planes were set at their focal planes. The scale of the simulation and experimental results are the same. For (a), dashed circles indicate the locations of the second and third pressure valleys for both simulation and experimental results. For (b), white dashed circles indicate the second and fourth pressure valleys.

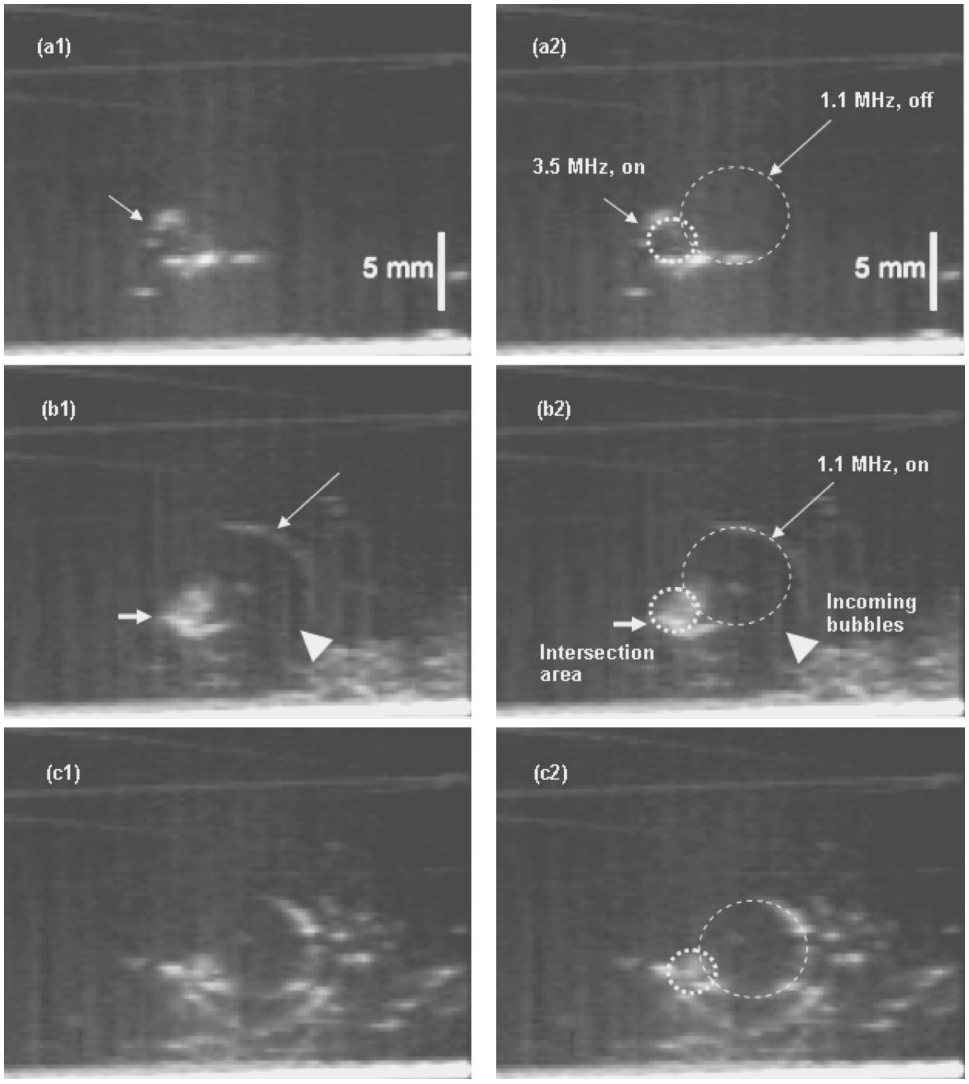


FIG. 10. The combined effect of the 1.1 and 3.5 MHz HIFU transducers is shown. The left column (a1, b1, and c1) is the original B-mode images taken at the confocal plane of the two transducers. Locations of the first rings of both transducers are shown by thick (3.5 MHz) and thin (1.1 MHz) dashed rings. When the 3.5 MHz HIFU transducer is on but the 1.1 one is off, the first ring of its beam pattern is clearly seen (thin arrow in a1 and a2). After turning on the 1.1 MHz transducer and supplying the focal area with shock wave-generated bubbles, the focus of the 3.5 MHz transducer filled with high-scattering signals, most likely bubbles (thick arrow in b1 and b2). The bubbles were generated and destroyed repeatedly at the focus of the 3.5 MHz beam. The moving direction of the shock wave-generated bubbles is indicated by the arrow heads (b1 and b2). The bubble-like signals at the focus of the 3.5 MHz transducer can sustain for up to a few minutes after supplying ''seed'' bubbles generated by shock waves once. The c1 and c2 plots were taken after about a minute after stopping supplying bubbles from shock waves.

bubbles by the shock wave transducer to the combined focus induced inertial cavitation, and the enhanced scattering was sustained for longer periods of time (up to a few minutes). In our case, the focal pressure of the 3.5 MHz transducer increased by summation with the pressure of the first sidelobe of the 1.1 MHz transducer, and exceeded the inertial cavitation threshold at 3.5 MHz. We believe that bubbles were generated, grew, and collapsed continuously at the intersecting area.

The schlieren imaging uses light to form images, which will not alter the sound field of the target transducer, and is pretty fast. However, the shlieren imaging method is basically a semiquantitative measurement of the pressure profile of the ultrasound beam. The bright and dark patterns projected on a screen can be used to determine the relative positions of the focus and sidelobes. Moreover, an absolute value of the beam size cannot be easily obtained. Furthermore, the schlieren imaging could be distorted easily at the presence of bubbles from inertial cavitation. The laser light diffracted while passing through moving bubbles and thus reduced the quality of the generated images. The new imaging method using bubbles can be used to perform measurement quantitatively. The obtained image quality even improves at the presence of bubbles from inertial cavitation.

The ''step'' behavior seen in Fig. 5 represented the true pressure profile around the HIFU focus. As seen in Fig. 11(a), the bright parts represent the pressure crests (peaks and saddles) while the dark parts are the pressure valleys. This background beam plot was created using the simulation equations described above. Bubbles tended to move to the low-pressure area and thus accumulated in the dark regions. The measured radii of the beam profile represented the inner most pressure valleys which bubbles could reach at different distances from the transducer surface. When the transducer's output was low, the pressure at the saddle part [A in Fig. 11(b)] was low enough to allow bubbles to climb over it and stayed in the first pressure valley, the lowest ''step'' (46–50 mm). The pressure gradually increased while the distance to the transducer increased, and prevented bubbles from further climbing over. Bubbles stayed in the second pressure valley and formed a higher step (B, 50–54 mm). Ambiguity occurred at the transition between two steps. That is, bubbles can stay either in the lower or the higher pressure valley. Figure 11(b) shows the relative pressure amplitudes of the first pressure valley, the pressure crest, and the second pressure valley. However, when the output pressure was high enough [Fig. 5(b)], bubbles were more and more difficult to

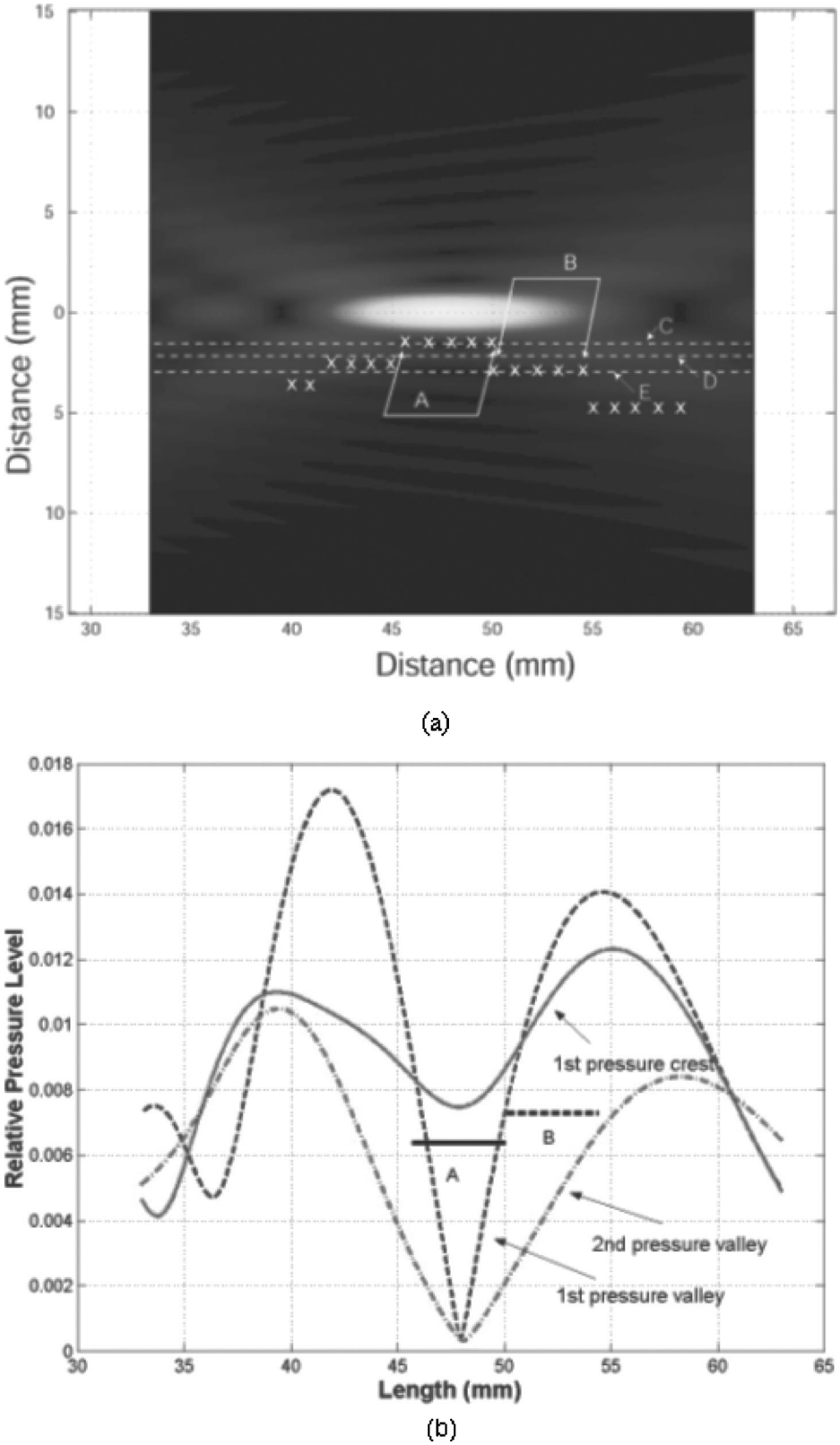


FIG. 11. (a) Simulated pressure profile of the 1.1 MHz HIFU transducer near the transducer focus. ''X'' represents the positions where focal sizes were measured (see Fig. 5). ''A'' is the lowest pressure step at position 46 to 50 mm. The pressure ''crest (peak and saddle)'' outside A is low and bubbles can climb over and enter ''A'' to form the first ring. ''B'' is where the second pressure ''step'' locates (50 to 54 mm). The detailed pressure profiles of slice C (through step ''A''), D (through first pressure crest), and E (through step ''B'') are shown in (b).

enter the inner pressure valleys and the measured radii thus increased rapidly.

Our observations indicate that shock waves can be used to provide nuclei to reduce the cavitation threshold and to locally induce inertial cavitation activity. The focal size of a shock wave is small ($1.7\times1.7\times6.7$ mm), and thus a large amount of energy can be directed to a small region *in vivo*. The negative pressure of the shock wave was sufficiently low to induce cavitation and create bubbles in the intrahepatic vessels and bile.[9,10] It is also possible for bubbles to be generated in loose tissue such as liver parenchyma. The shock wave-generated bubbles can be used to provide bubble seeds for further cavitation activity, and thus may be used to facilitate ultrasound-related bioeffects, such as drug delivery or

gene transfection, when used with a therapeutic ultrasound transducer.

ACKNOWLEDGMENTS

This research was supported by grants from the National Science Council No. 92-2312-B-002-019-, and the National Taiwan University Hospital No. 93N011.

[1] S. Vaezy, V. Y. Fujimoto, C. Walker, R. W. Martin, E. Y. Chi, and L. A. Crum, ''Treatment of uterine fibroid tumors in a nude mouse model using high-intensity focused ultrasound,'' Am. J. Obstet. Gynecol. **183**, 6–11 (2000).

[2] G. ter Haar, D. Sinnett, and I. Rivens, ''High intensity focused ultrasound—a surgical technique for the treatment of discrete liver tumours,'' Phys. Med. Biol. **34**, 1743–1750 (1989).

[3] F. Wu, Z. B. Wang, W. Z. Chen, J. Z. Zou, J. Bai, H. Zhu, K. Q. Li, F. L. Xie, C. B. Jin, H. B. Su, and G. W. Gao, ''High intensity focused ultrasound for extracorporeal treatment of solid carcinomas: four-year Chinese clinical experience,'' *Proceedings of the International Symposium of Therapeutic Ultrasound*, 2002, pp. 34–43.

[4] R. Yang, C. R. Reilly, F. J. Rescorla, P. R. Faught, N. T. Sanghvi, F. J. Fry, T. D. Franklin, Jr., L. Lumeng, and J. L. Grosfeld, ''High-intensity focused ultrasound in the treatment of experimental liver cancer,'' Arch. Surg. (Chicago) **126**, 1002–1009 (1991).

[5] J. Y. Chapelon, M. Ribault, F. Vernier, R. Souchon, and A. Gelet, ''Treatment of localised prostate cancer with transrectal high intensity focused ultrasound,'' Eur. J. Ultrasound **9**, 31–38 (1999).

[6] U. Schatzle, T. Reuner, J. Jenne, and A. Heilingbrunner, ''Quality assurance tools for therapeutic ultrasound,'' Ultrasonics **36**, 679–682 (1998).

[7] H. T. O'Nell, ''Theory of focusing radiators,'' J. Acoust. Soc. Am. **21**, 516–526 (1949).

[8] M. Blomley, T. Albrecht, D. Cosgrove, V. Jayaram, J. Butler-Barnes, and R. Eckersley, ''Stimulated acoustic emission in liver parenchyma with Levovist,'' Lancet **351**, 568 (1998).

[9] M. Delius, R. Denk, C. Berding, H. G. Liebich, M. Jordan, and W. Brendel, ''Biological effects of shock waves: cavitation by shock waves in piglet liver,'' Ultrasound Med. Biol. **16**, 467–472 (1990).

[10] L. E. Forer, W. J. Davros, J. Goldberg, F. al Kawas, B. S. Garra, W. Hayes, and R. K. Zeman, ''Hepatic cavitation. A marker of transient hepatocellular injury during biliary lithotripsy,'' Dig. Dis. Sci. **37**, 1510–1516 (1992).

Frequency domain expressions for the estimation of time-averaged acoustic energy density

Ben S. Cazzolato[a] and Justin Ghan[b]
School of Mechanical Engineering, The University of Adelaide, SA 5005, Australia

(Received 7 September 2004; revised 14 February 2005; accepted 14 March 2005)

This paper builds on earlier work by the same authors to derive expressions for the time-averaged acoustic energy density in the frequency domain using the auto- and cross-spectral densities of multiple microphone elements. Expressions for the most common three-dimensional geometric arrangements are derived. Simplified expressions for use with two channel spectrum analysers are also presented. [DOI: 10.1121/1.1904505]

PACS numbers: 43.58.-e, 43.60.Qv [AJZ] Pages: 3750–3756

I. INTRODUCTION

In several articles[1–3] an expression for the acoustic energy density spectral density along a single axis using the two-microphone technique has been derived. This frequency domain expression enables the calculation of acoustic energy density spectral density using the autospectra and cross-spectrum of the two microphone signals. The advantage of this approach, as opposed to the alternative time-domain approach[4] of first calculating the instantaneous mean pressure and particle velocity, is that the energy density may be measured with a FFT based two-channel spectrum analyzer with no additional circuitry apart from the two microphone signals.

In this paper, the acoustic energy density estimate using spectral methods will be extended to three dimensions. Several geometric configurations and spectral density estimation formulations are discussed. Expressions involving the minimum number of terms necessary to estimate the energy density spectral density are also derived to facilitate the calculation of energy density in 3D sound fields using a two-channel spectrum analyser.

II. PHYSICAL 3D ENERGY DENSITY CONFIGURATIONS

Since the measurement of 3D acoustic intensity requires the same set of measurements as acoustic energy density, namely acoustic pressure and the three orthogonal particle velocities, then any sensor capable of measuring one quantity can also be used to measure the other. There exists several different geometrical arrangements for which 3D acoustic intensity and energy density may be measured.[5–8] This paper will loosely focus on the two most common arrangements found in commercially available 3D sound intensity probes. These are the four-microphone tetrahedral arrangement like those found in the Ono Sokki Tetra-phone MI-6420 [Fig. 1(a)] and the six-microphone arrangement like those of the Brüel & Kjær Type 5356 Intensity Probe [Fig. 1(b)] or the GRAS Vector Intensity Probe [Fig. 1(c)].

A number of variants of the four-microphone sensor exist, the specific details of which are discussed in detail in the following sections. A slight variant of the six-microphone sensor is the seven-microphone sensor.[9] Both the six and seven-microphone sensors use three pairs of opposed closely spaced microphones to calculate the three orthogonal particle velocities. However unlike its six-microphone sibling, which must estimate the pressure at the center of the sensor using a finite sum, the seven-microphone sensor uses the seventh microphone at its geometric center to directly measure the pressure. This is the most accurate of all the geometries[7] and consequently has been used to accurately measure 3D impulse response functions for surround sound systems.[9]

Parkins *et al.*[8] investigated a six-microphone energy density with the microphones mounted on the surface of a rigid sphere [rather than the open configurations shown in Figs. 1(b) and 1(c)]. This arrangement was employed in an attempt to use the effects of diffraction caused by the sphere[10] to produce more favorable bias conditions, and in doing so reduce the inherent errors. The effect of acoustic diffraction around the rigid sphere caused the effective microphone spacing to be 50% greater than the actual microphone spacing. When calculating the particle velocity estimate when using the finite difference approximation, the effective microphone spacing should be used, rather that the physical spacing.

III. ANALYTICAL DERIVATION

The analytical derivation in this paper closely follows the derivation by Ghan *et al.*[3] The instantaneous acoustic energy density, $E_D(t)$, at a point is defined as the sum of the acoustic potential energy density and the acoustic kinetic energy density at that point, given by[6]

$$E_D(t)=\frac{p^2(t)}{2\rho c^2}+\frac{\rho|\mathbf{v}(t)|^2}{2}, \tag{1}$$

where $p(t)$ is the instantaneous pressure and $|\mathbf{v}(t)|$ is the magnitude of the total instantaneous particle velocity, c is the speed of sound, and ρ is the density of the fluid. In practice

[a]Author to whom correspondence should be addressed. Phone: +61 8 8303 5449; Fax: +61 8 8303 4367; Electronic mail: benjamin.cazzolato@adelaide.edu.au
[b]Electronic mail: justin.ghan@adelaide.edu.au

 0001-4966/2005/117(6)/3750/7/$22.50

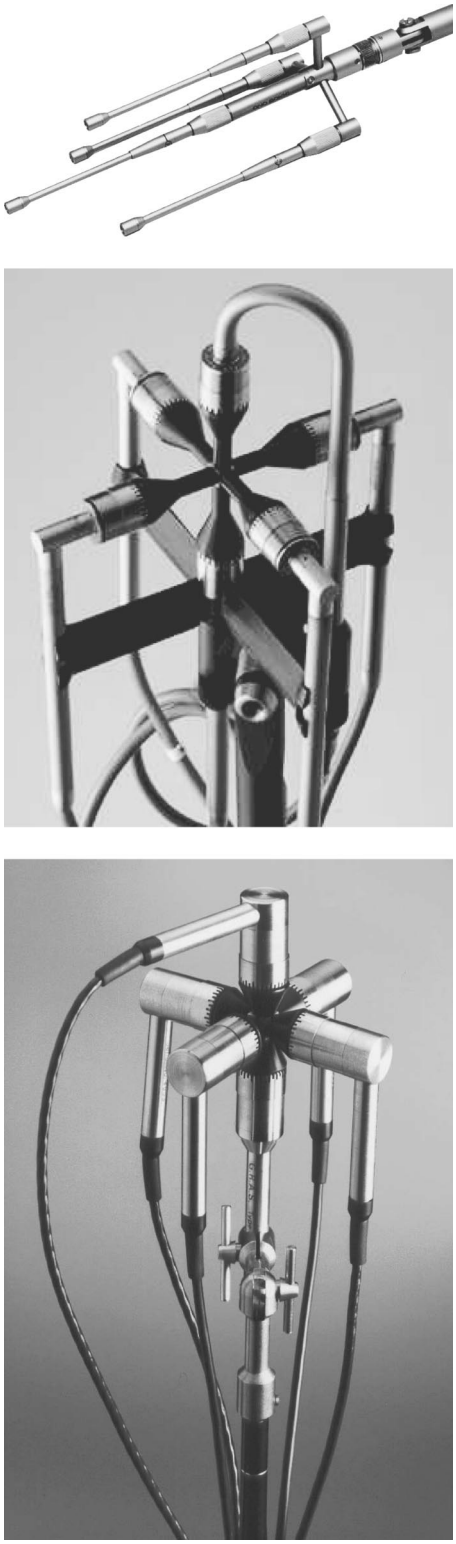


FIG. 1. Commercially available sound intensity sensors suitable for 3D energy density measurement. (a) Ono Sokki Tetra-phone MI-6420. Photo courtesy of Ono Sokki; (b) B&K Type 5356. Photo courtesy of Brüel and Kjær; (c) GRAS Vector Intensity Probe. Photo courtesy of GRAS.

the total particle velocity is estimated using the three orthogonal terms $[\mathbf{v}(t)=\mathbf{i}v_x(t)+\mathbf{j}v_y(t)+\mathbf{k}v_z(t)]$, therefore Eq. (1) can be rewritten as

$$E_D(t)=\frac{p^2(t)}{2\rho c^2}+\sum_{i=x,y,z}\frac{\rho v_i^2(t)}{2}. \tag{2}$$

The pressure is typically estimated using an average of all the microphone elements, thus for an sensor containing n microphones the pressure estimate is

$$p(t)\approx\frac{1}{n}\sum_{i=1}^{n}p_i(t). \tag{3}$$

The particle velocity can be estimated through a finite difference approximation using two microphones located on one of the three orthogonal axes, thus for a microphone separation distance of $2h$, the particle velocity components are approximated by[6]

$$v_i(t)\approx\frac{1}{2\rho h}\int_{-\infty}^{t}[p_{i_1}(\tau)-p_{i_2}(\tau)]\mathrm{d}\tau. \tag{4}$$

In the following sections, frequency domain expressions are derived for the time-averaged acoustic energy density estimate for a number of 3D sensor geometries. The expressions are based on single-sided spectral densities so they may be directly used with real-time spectrum analysers which usually calculate single-sided spectra. All expressions have been validated using the same technique employed by Ghan *et al.*[3]

A. Multichannel expressions

Frequency domain expressions for the most common 3D geometries are derived in the following sections. In the case where a multichannel spectrum analyzer (such as a Brüel and Kjær pulse system) is available, it is possible to calculate the expressions in real time. Evaluation of the expressions using a two-channel spectrum analyser can be made by recording all auto and cross-spectra, and then post processing the spectra in a software package such as Matlab or Excel.

1. Cubic four-microphone sensor with pressure averaging

The following analysis is for the four-microphone configuration as shown in Fig. 2(a). This particular geometric configuration is known as a ''cubic'' arrangement since the microphones lie in the corner of a cube, where the origin microphone (marked 1) is in one corner of the cube and the three other microphones (marked 2, 3, and 4) are located on the corresponding vertices of the cube, each forming one of the 3 orthogonal axes (marked x, y, and z, respectively).

The specifics of this particular four-microphone sensor when used for energy density sensing is discussed in detail by Cazzolato and Hansen.[7] This energy density sensor arrangement has been widely applied to the characterization of reverberant sound fields.[4,11] A frequency domain expression for the sound intensity spectral density from this arrangement has been derived,[12] although to the authors' knowledge, no such equivalent expressions exists for energy density.

The pressure estimate is the mean of the pressures measured by all four microphones. The particle velocities are calculated using Eq. (4). The distance from the ''origin'' microphone (microphone 1) to the other microphones was $2h$. It should be noted that this particular sensor does not have a unique acoustic center.[7,12,13] This means that the positions in which the particle velocities are estimated are not coincident with each other or the location of the pressure estimate. The

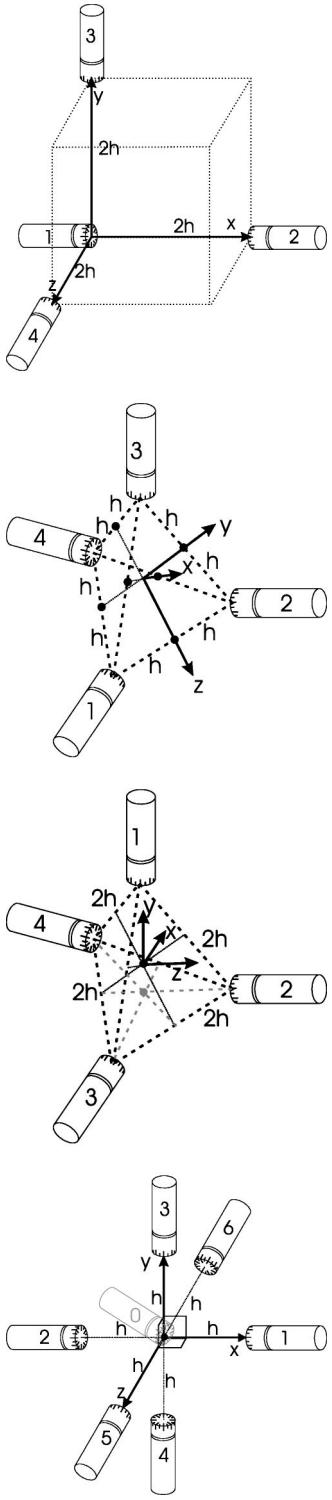


FIG. 2. Multimicrophone three-axis energy density sensors. (a) Cubic four-microphone sensor. (b) Four-microphone regular tetrahedon sensor. Small black circles represent the virtual microphone locations. (c) Four-microphone Ono Sokki sensor. (d) Six-microphone and seven-microphone sensor.

alternative four-microphone arrangements shown in Figs. 2(b) and (c) and discussed in more detail in Secs. III A 2 and III A 3 have a unique acoustic center which is located at the center of gravity of the microphones.

The total time-averaged acoustic energy density is given by[3]

$$\overline{E_D} = \lim_{T\to\infty} \frac{1}{T} E\left[\int_0^T \left[\frac{1}{2\rho c^2} \cdot p^2(t) + \frac{\rho}{2} \cdot |\mathbf{v}(t)|^2 \right] \mathrm{d}t \right], \quad (5)$$

where the operator E is the expectation and T is the record length.

Using Parseval's theorem[14] it can be shown that the double-sided time-averaged acoustic energy density spectral density is,

$$\overline{E_D}(\omega) = \lim_{T\to\infty} \frac{1}{T} E\left[\frac{1}{2\rho c^2} \cdot |P(\omega,T)|^2 + \frac{\rho}{2} \cdot |V_x(\omega,T)|^2 + \frac{\rho}{2} \cdot |V_y(\omega,T)|^2 + \frac{\rho}{2} \cdot |V_z(\omega,T)|^2 \right], \quad (6)$$

where $P(\omega,T)$ and $V_i(\omega,T)$ are the Fourier transforms of the pressure and velocity estimates. From Eqs. (3) and (4),

$$P(\omega,T) \approx \frac{P_1(\omega,T) + P_2(\omega,T) + P_3(\omega,T) + P_4(\omega,T)}{4}, \quad (7)$$

$$V_x(\omega,T) \approx \frac{1}{2\rho h} \cdot \frac{P_1(\omega,T) - P_2(\omega,T)}{j\omega}, \quad (8)$$

$$V_y(\omega,T) \approx \frac{1}{2\rho h} \cdot \frac{P_1(\omega,T) - P_3(\omega,T)}{j\omega}, \quad (9)$$

$$V_z(\omega,T) \approx \frac{1}{2\rho h} \cdot \frac{P_1(\omega,T) - P_4(\omega,T)}{j\omega}. \quad (10)$$

Equation (6) is valid for all sensor configurations, however the expressions for the pressure and particle velocity differ for each sensor. Evaluating the expectation operator in Eq. (6) it can be shown that the single-sided time-averaged energy density spectral density estimate for the four-microphone sensor with pressure averaging is given by

$$\begin{aligned} \overline{E_D}1\text{-sided}(\omega)_{\overline{4}} \approx & \left(\frac{1}{32\rho c^2} + \frac{3}{8\rho\omega^2 h^2} \right)(G_{11}) \\ & + \left(\frac{1}{32\rho c^2} + \frac{1}{8\rho\omega^2 h^2} \right)(G_{22} + G_{33} + G_{44}) \\ & + \left(\frac{1}{32\rho c^2} - \frac{1}{8\rho\omega^2 h^2} \right)(2\,\mathrm{Re}[G_{12}] \\ & + 2\,\mathrm{Re}[G_{13}] + 2\,\mathrm{Re}[G_{14}]) + \left(\frac{1}{32\rho c^2} \right) \\ & \times (2\,\mathrm{Re}[G_{23}] + 2\,\mathrm{Re}[G_{24}] + 2\,\mathrm{Re}[G_{34}]), \end{aligned} \quad (11)$$

where G_{xx} is the single-sided auto-spectral density function

for the pressure signal measured by microphone x, and $\mathrm{Re}[G_{xy}]$ is the real part of the single-sided cross-spectral density function for microphone signals x and y. The subscript 4 represents the pressure estimate is obtained by averaging all four microphone signals.

2. Regular tetrahedral four-microphone sensor with pressure averaging

Santos *et al.*[13] described a four-microphone arrangement in which the ''acoustic center'' is the same as the geometric center of the four microphones [see Fig. 2(b)]. Their design used microphones mounted on the corners of a regular tetrahedron. They combined pairs of the microphone signals to estimate the pressure at the midpoint of each edge of the tetrahedron. These six ''virtual points'' are then used to calculate the pressure gradient at the geometric center of the sensor.

The pressure estimate is the mean pressure from all four microphones. The double-sided time-averaged acoustic energy density spectral density is given by Eq. (6) where,[13]

$$P(\omega,T)\approx\frac{P_1(\omega,T)+P_2(\omega,T)+P_3(\omega,T)+P_4(\omega,T)}{4}, \tag{12}$$

$$V_x(\omega,T)\approx\frac{1}{\sqrt{2}\rho h j\omega}\cdot\left(\frac{P_1(\omega,T)+P_3(\omega,T)}{2}-\frac{P_2(\omega,T)+P_4(\omega,T)}{2}\right), \tag{13}$$

$$V_y(\omega,T)\approx\frac{1}{\sqrt{2}\rho h j\omega}\cdot\left(\frac{P_1(\omega,T)+P_4(\omega,T)}{2}-\frac{P_2(\omega,T)+P_3(\omega,T)}{2}\right), \tag{14}$$

$$V_z(\omega,T)\approx\frac{1}{\sqrt{2}\rho h j\omega}\cdot\left(\frac{P_1(\omega,T)+P_2(\omega,T)}{2}-\frac{P_3(\omega,T)+P_4(\omega,T)}{2}\right). \tag{15}$$

It can be shown that the single-sided time-averaged energy density spectral density estimate for the regular tetrahedron four-microphone sensor with pressure averaging is given by

$$\begin{aligned}\overline{E_{\mathrm{D}}}1\text{-sided}(\omega)_{\mathrm{Tetrahedron}}&\approx\left(\frac{1}{32\rho c^2}+\frac{3}{16\rho\omega^2h^2}\right)(G_{11}+G_{22}+G_{33}+G_{44})\\&+\left(\frac{1}{32\rho c^2}-\frac{1}{16\rho\omega^2h^2}\right)(2\,\mathrm{Re}[G_{12}]+2\,\mathrm{Re}[G_{13}]\\&+2\,\mathrm{Re}[G_{14}])+\left(\frac{1}{32\rho c^2}-\frac{1}{16\rho\omega^2h^2}\right)(2\,\mathrm{Re}[G_{23}]\\&+2\,\mathrm{Re}[G_{24}]+2\,\mathrm{Re}[G_{34}]).\end{aligned} \tag{16}$$

3. Ono Sokki four-microphone sensor with pressure averaging

The Ono Sokki four-microphone sensor[15,16] shown in Fig. 1(a) also has the microphones arranged in a regular tetrahedron. The weighting of the microphone signal differs from that employed by Santos *et al.*[13] resulting in a different alignment of the three measurement axes [see Fig. 2(c)].

Using the approach employed by the Ono Sokki corporation when processing the data, the double-sided time-averaged acoustic energy density spectral density is given by Eq. (6) where[16]

$$P(\omega,T)\approx\frac{P_1(\omega,T)+P_2(\omega,T)+P_3(\omega,T)+P_4(\omega,T)}{4}, \tag{17}$$

$$V_x(\omega,T)\approx\frac{1}{2\rho h j\omega}\cdot(P_3(\omega,T)-P_2(\omega,T)), \tag{18}$$

$$V_y(\omega,T)\approx\frac{1}{2\sqrt{3}\rho h j\omega}\cdot(2P_1(\omega,T)-P_2(\omega,T)-P_3(\omega,T)), \tag{19}$$

$$V_z(\omega,T)\approx\frac{1}{2\sqrt{6}\rho h j\omega}\cdot(3P_4(\omega,T)-P_1(\omega,T)-P_2(\omega,T)-P_3(\omega,T)). \tag{20}$$

It can be shown that the single-sided time-averaged energy density spectral density estimate for the regular tetrahedron four-microphone Ono Sokki sensor with pressure averaging is identical to that of the regular tetrahedral sensor given by Eq. (16). This is to be expected since the pressure and total velocity estimate is the same for both cases, only the orientation with respect to the coordinate system has changed.

4. Six-microphone sensor with pressure averaging

Figure 2(d) shows a sensor containing six microphones, with three pairs forming three orthogonal axes. The mean of all six microphones signals provides the pressure estimate. The double-sided time-averaged acoustic energy density spectral density is given by Eq. (6) where,

$$P(\omega,T)\approx\frac{P_1(\omega,T)+P_2(\omega,T)+P_3(\omega,T)+P_4(\omega,T)+P_5(\omega,T)+P_6(\omega,T)}{6}, \tag{21}$$

$$V_x(\omega,T)\approx\frac{1}{2\rho h}\cdot\frac{P_1(\omega,T)-P_2(\omega,T)}{j\omega}, \tag{22}$$

$$V_y(\omega,T)\approx\frac{1}{2\rho h}\cdot\frac{P_3(\omega,T)-P_4(\omega,T)}{j\omega}, \tag{23}$$

$$V_z(\omega,T)\approx\frac{1}{2\rho h}\cdot\frac{P_5(\omega,T)-P_6(\omega,T)}{j\omega}. \tag{24}$$

Therefore it can be shown that the single-sided time-averaged energy density spectral density estimate for the six-microphone sensor with pressure averaging is given by

$$\begin{aligned}\overline{E_{\mathrm{D}}}1\text{-sided}(\omega)_{\overline{6}}\approx&\left(\frac{1}{72\rho c^2}+\frac{1}{8\rho\omega^2h^2}\right)(G_{11}+G_{22}+G_{33}\\&+G_{44}+G_{55}+G_{66})+\left(\frac{1}{72\rho c^2}-\frac{1}{8\rho\omega^2h^2}\right)\\&\times(2\,\mathrm{Re}[G_{12}]+2\,\mathrm{Re}[G_{34}]+2\,\mathrm{Re}[G_{56}])\\&+\left(\frac{1}{72\rho c^2}\right)\left(\sum_{i=3}^{6}(2\,\mathrm{Re}[G_{1i}]+2\,\mathrm{Re}[G_{2i}])\right.\\&\left.+\sum_{i=5}^{6}(2\,\mathrm{Re}[G_{3i}]+2\,\mathrm{Re}[G_{4i}])\right).\end{aligned} \tag{25}$$

If the six microphones are mounted in a rigid sphere as described by Parkins *et al.*,[8] then the substitution $h=\frac{3}{2}r$ should be made, where r is radius of the sphere.

5. Seven-microphone sensor

With reference to the seven-microphone configuration in Fig. 2(d), the additional "origin" microphone (p_0), located at the geometric center of the six-microphone sensor, is used to measure the pressure. By measuring the pressure directly, the error associated with the finite-sum is avoided.

The single-sided time-averaged energy density spectral density estimate for the seven-microphone sensor is given by

$$\begin{aligned}\overline{E_{\mathrm{D}}}1\text{-sided}(\omega)_7\approx&\left(\frac{1}{2\rho c^2}\right)(G_{00})+\left(\frac{1}{8\rho\omega^2h^2}\right)(G_{11}+G_{22}\\&+G_{33}+G_{44}+G_{55}+G_{66})-\left(\frac{1}{8\rho\omega^2h^2}\right)\\&\times(2\,\mathrm{Re}[G_{12}]+2\,\mathrm{Re}[G_{34}]\\&+2\,\mathrm{Re}[G_{56}]),\end{aligned} \tag{26}$$

where G_{00} is the autospectral density function of the origin microphone signal. Note the significant reduction in terms compared to the six-microphone sensor.

B. Reduced order expressions

Although it is desirable to calculate all the necessary auto- and cross-spectra simultaneously on a multichannel spectrum analyzer, a two-channel spectrum analyzer (such as the HP 35665A) may be still be used. The number of two-channel measurements required is a function of the number of microphones and the particular spectral density formulation. To calculate the energy density spectral density estimates using Eqs. (11) and (16) requires 6 individual two-channel measurements and Eqs. (25) and (26) would require 15 and 4 individual two-channel measurements, respectively.

Clearly the large number of two-channel measurements for all but the seven-microphone sensor would require a significant effort to obtain experimentally. It is therefore desirable to have expressions requiring fewer measurements. The minimum number of separate measurements is limited by the need for the cross-spectral density estimates used by the finite-difference expression in the kinetic energy density estimate. In the following sections a number of approximations using less terms are derived to facilitate calculation using two-channel spectrum analysers.

1. Cubic four-microphone sensor using origin microphone

An alternative formulation[7,17] for the four-microphone sensor is to use the pressure at the "origin microphone" of the cubic arrangement as the pressure estimate rather than the mean pressure sensed by the four microphones as used previously. The particle velocities may be calculated as before using Eqs. (8)–(10). For this arrangement the single-sided time-averaged energy density spectral density estimate is given by

$$\begin{aligned}\overline{E_{\mathrm{D}}}1\text{-sided}(\omega)_{4_0}\approx&\left(\frac{1}{2\rho c^2}+\frac{3}{8\rho\omega^2h^2}\right)(G_{11})\\&+\left(\frac{1}{8\rho\omega^2h^2}\right)(G_{22}+G_{33}+G_{44})\\&+\left(\frac{1}{8\rho\omega^2h^2}\right)(-2\,\mathrm{Re}[G_{12}]\\&-2\,\mathrm{Re}[G_{13}]-2\,\mathrm{Re}[G_{14}]).\end{aligned} \tag{27}$$

This formulation only requires three individual sets of measurements as compared to six for Eqs. (11) and (16).

2. Six-microphone sensor: Reduced order

An approach commonly employed for 3D sound intensity measurements is to use three 1D probes.[5] This approach was used by Parkins *et al.*[8] when calculating the energy density for a six-microphone sensor. They used the mean of three two-channel pressure estimates (obtained from the finite pressure sum of each of the three orthogonal axes) to provide an overall pressure estimate. Thus, when using this approach the square of the pressure magnitude in Eq. (6) is given by

$$|P(\omega,T)|^2 \approx \frac{1}{3}\left[\left|\frac{P_1(\omega,T)+P_2(\omega,T)}{2}\right|^2 + \left|\frac{P_3(\omega,T)+P_4(\omega,T)}{2}\right|^2 + \left|\frac{P_5(\omega,T)+P_6(\omega,T)}{2}\right|^2\right]. \tag{28}$$

For this particular two-channel approximation, the single-sided time-averaged energy density spectral density estimate is given by

$$\overline{E_{\mathrm{D}}}1\text{-sided}(\omega)_{6\bar{3}} \approx \left(\frac{1}{24\rho c^2}+\frac{1}{8\rho\omega^2h^2}\right)(G_{11}+G_{22}+G_{33}+G_{44}+G_{55}+G_{66}) + \left(\frac{1}{24\rho c^2}-\frac{1}{8\rho\omega^2h^2}\right) \times (2\,\mathrm{Re}[G_{12}]+2\,\mathrm{Re}[G_{34}]+2\,\mathrm{Re}[G_{56}]). \tag{29}$$

This formulation also only requires three individual sets of two-channel measurements compared to 15 two-channel measurements needed when using Eq. (25). This probably explains why this particular approach was used by Parkins *et al.*[8] when estimating the 3D energy density. It should also be noted that the expression derived by Parkins *et al.*[8] differs from Eq. (29).

3. Cubic four-microphone sensor: Reduced order

The method used in Sec. III B 2 has been applied to a cubic four-microphone sensor to derive an expression for the energy density using a reduced number of two-channel measurements. Thus the square of the pressure magnitude in Eq. (6) is given by

$$|P(\omega,T)|^2 \approx \frac{1}{3}\left[\left|\frac{P_1(\omega,T)+P_2(\omega,T)}{2}\right|^2 + \left|\frac{P_1(\omega,T)+P_3(\omega,T)}{2}\right|^2 + \left|\frac{P_1(\omega,T)+P_4(\omega,T)}{2}\right|^2\right]. \tag{30}$$

For this four-microphone sensor approximation, the single-sided time-averaged energy density spectral density estimate is given by

$$\overline{E_{\mathrm{D}}}1\text{-sided}(\omega)_{4\bar{3}} \approx \left(\frac{1}{8\rho c^2}+\frac{3}{8\rho\omega^2h^2}\right)(G_{11}) + \left(\frac{1}{24\rho c^2}+\frac{1}{8\rho\omega^2h^2}\right)(G_{22}+G_{33}+G_{44}) + \left(\frac{1}{24\rho c^2}-\frac{1}{8\rho\omega^2h^2}\right)(2\,\mathrm{Re}[G_{12}]+2\,\mathrm{Re}[G_{13}]+2\,\mathrm{Re}[G_{14}]). \tag{31}$$

This formulation only requires three individual sets of measurements as compared to six for Eqs. (11) and (16). It is interesting to note that Eq. (31) contains the same auto and cross-spectral density terms as Eq. (27), however the weights of the individual terms differ, therefore likely producing different results under certain conditions.

The reason for the differing weights in the cubic four-microphone expressions given by Eqs. (11), (27), and (31) is the location at where the pressure estimate is made. For Eq. (11), the geometric center of the four microphones is used, which is located a distance of $h\sqrt{3}/2$ from the origin microphone along a line normal to the plane intersecting the other three points. The pressure estimate using Eq. (27) is obviously made at the origin microphone. The estimate of the pressure using Eq. (31) is made at a distance of $h/\sqrt{3}$ from the origin microphone, which is halfway between the origin microphone and the plane intersecting the other three points.

Each component of the velocity is estimated halfway between the origin microphone and one of the axis microphones. Given that the distance between the location of the pressure estimate and velocity estimates using Eq. (31) is slightly less than that obtained by using Eq. (11), namely $h\sqrt{2}/\sqrt{3}=0.8165h$ compared to $h\sqrt{3}/2=0.8660h$, it could be argued that the former is the most accurate formulation.

IV. CONCLUSIONS

Several expressions for the time-averaged acoustic energy density estimate have been derived using the auto- and cross-spectral densities between several closely spaced microphones. Expressions containing a reduced number of terms were also derived to facilitate the calculation of energy density using two-channel spectrum analyzers.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support for this work provided by the Australian Research Council.

[1] G. W. Elko, "Simultaneous measurement of the complex acoustic intensity and the acoustic energy density," in Proceedings of the 2nd International Congress on Acoustic Intensity, 1985, pp. 69–78.

[2] T. E. Vigran, "Acoustic intensity—energy density ratio: An index for detecting deviations from ideal field conditions," J. Sound Vib. **127**, 343–351 (1988).

[3] J. Ghan, B. S. Cazzolato, and S. D. Snyder, "Expression for the estimation of time-averaged acoustic energy density using the two-microphone method," J. Acoust. Soc. Am. **113**, 2404–2407 (2003).

[4] J. A. Moryl and E. L. Hixson, "A total acoustic energy density sensor with

applications to energy density measurement in a reverberation room,'' in *Proceedings of Inter-Noise 87*, 1987, pp. 1195–1198.
[5] G. Rasmussen, ''Measurement of vector fields,'' in *Proceedings of the 2nd International Congress on Acoustic Intensity*, 1985, pp. 53–58.
[6] F. Fahy, *Sound Intensity*, 2nd ed. (E&FN Spon, London, 1995).
[7] B. S. Cazzolato and C. H. Hansen, ''Errors arising from three-dimensional acoustic energy density sensing in one-dimensional sound fields,'' J. Sound Vib. **236**, 375–400 (2000).
[8] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, ''Error analysis of a practical energy density sensor,'' J. Acoust. Soc. Am. **108**, 211–222 (2000).
[9] A. Farina and L. Tronchin, ''3D impulse response measurements on S. Maria del Fiore Church,'' Florence, Italy, in *Proceedings of ICA98—International Conference on Acoustics*, Seattle, Washington State, 1998, pp. 26–30.
[10] G. W. Elko, ''An acoustic vector-field probe with calculable obstacle bias,'' in *Proceedings of Noise-Con 91*, 1991, pp. 525–532.
[11] M. H. W. Budhiantho and E. L. Hixson, ''A proposal to adopt Maxwell distribution as a measure of acoustic field diffusness in a reverberation room,'' J. Acoust. Soc. Am. **105**, 935 (1998).
[12] J. Vandenhout, P. Sas, and R. Snoeys, ''Measurement, accuracy and interpretation of real and imaginary intensity patterns in the near field of complex radiators,'' in *Proceedings of the 2nd International Congress on Acoustic Intensity*, 1985, pp. 121–128.
[13] I. M. C. Santos, C. C. Rodrigues, and J. L. Bento Coelho, ''Measuring the three-dimensional acoustic intensity vector with a four-microphone probe,'' in *Proceedings of Inter-Noise 89*, 1989, pp. 965–968.
[14] J. S. Bendat and A. G. Piersol, *Random Data—Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986).
[15] K. Yamaguchi, K. Hori, T. Tananka, and M. Anzai, ''The development of instruments for the measurement of sound intensity using 3-dimensional microphone probe,'' in *Proceedings on Inter-Noise 94*, 1994, pp. 1967–1970.
[16] M. Suzuki, H. Anzai, S. Oguro, and T. Ono, ''Performance evaluation of a three dimensional intensity probe,'' J. Acoust. Soc. Jpn. (E) **16**, 233–238 (1995).
[17] B. S. Cazzolato and C. H. Hansen, ''Active control of enclosed sound fields using three-axis energy density sensors: Rigid walled enclosures,'' International Journal of Acoustics and Vibration **8**, 39–51 (2003).

Adaptive instant record signals applied to detection with time reversal operator decomposition

Thomas Folegot[a)]
Laboratoire Ondes et Acoustique, Université Denis Diderot, UMR CNRS 7587, ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 05, France and Atlantide, Marine Science and Technology Department, Technopôle Brest-Iroise, CS23866, 29238 Brest Cedex 3, France

Julien de Rosny, Claire Prada, and Mathias Fink
Laboratoire Ondes et Acoustique, Université Denis Diderot, UMR CNRS 7587, ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 27 November 2003; revised 15 March 2005; accepted 26 March 2005)

Time reversal arrays are becoming common tools whether for detection or tomography. These applications require the measurement of the response from the array to one or several receivers. The most natural way to record the impulse responses for several sources is to generate pulses successively from each emitting point and record simultaneously the signals from the receivers. However, this method is very time consuming or inefficient in terms of signal-to-noise ratio. To overcome this limitation quasi-orthogonal pseudonoise signals like Kasami sequences can be used. For guided wave propagation, a very high degree of orthogonality between the signal is necessary to allow an accurate measure of the whole multipath structure of the transfer function. Hence, in this work, we propose a new family of pseudo-orthogonal signals that is adapted to the environment and more specifically, to highly dispersive media. These adaptive instant records signals are used experimentally to detect targets using the time reversal operator decomposition method. The accuracy of the 15×15 transfer functions acquired simultaneously, and therefore the detection capability, are demonstrated in an experimental ultrasonic waveguide as a small-scale model of shallow water propagation including bottom absorption and reverberation. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1914204]

PACS numbers: 43.60.Ek, 43.60.Fg, 43.30.Gv, 43.20.Fn [DRD] Pages: 3757–3765

I. INTRODUCTION

Arrays of acoustic sources for underwater application have undergone tremendous development in the past several years. Such arrays are becoming common tools for detection, tomography, and communication. These applications require the measurement of the response from the array elements to one or several receivers and usually assume that the propagating medium is stationary. Unfortunately the temporal variability of the water column due to oceanographic phenomena, surface waves, and displacement of the transducers affects the response function so that these fluctuations can be significant especially in shallow water.

This problem has been encountered in time reversal experiments that have drawn considerable attention in underwater acoustics over the past 15 years.[1–12] Time-reversing acoustic technologies were proven to be effective for acoustic focusing under unknown acoustic environmental conditions. The experiments conducted in 1997 and in 1999 by the Marine Physical Laboratory and the NATO Undersea Research Center showed that a time reversal mirror can produce significant focusing at long distances in a 125 m deep-water channel. In the first experiment undertaken at 450 Hz, focusing was achieved out to a range of 30 km, and it was shown that probe-source pulses up to one week old can be refocused successfully. However, the second experiment done at the higher frequency of 3.5 kHz in the same environment[11] exhibited stability of less than 20 min. It would appear therefore that the effect of medium fluctuations is significant and severely limits the stability of focusing for such high frequency transmissions.

In this paper, we shall focus on the time reversal operator decomposition method (in French the D.O.R.T. method), which also involves arrays of transducers and is derived from the iterative time reversal. The efficiency of this method for separating the echo from two close targets through a complex medium has been shown in several papers.[13–16] The D.O.R.T. method has been applied to highly resolved detection in a water waveguide, and to the detection of a target close to the surface.[15] It has been demonstrated that the method provides the multipath structure of the propagation in shallow water as multiple reflections have been taken as an advantage to increase the resolution of the system. Numerical studies have since examined the possibility of using this method for target detection in underwater acoustics and small scale experimental studies showed how the method allows reverberation to be separated from target echo.[16] Nevertheless, the method requires a stationary waveguide and the long time necessary to acquire the array response signals is a major drawback. Even with parallel processed channels, for an array of N transducers, N acquisitions are required and the measurement is at least N times longer than the travel time.

Underwater communications systems working with multiple inputs and multiple outputs usually use pseudonoise

a)Electronic mail: folegot@nurc.nato.int

(PN) sequences. This family of orthogonal codes seems ideal to shorten the acquisition time by transmitting simultaneously information from several sources to a set of receivers. However, as shown in this paper, these signals do not have a sufficient orthogonality to measure the whole structure of the transfer function when the amplitude of the reflected acoustic travels expect to be about 25 dB lower than direct paths.

Here, we propose a method to overcome this difficulty using a set of orthogonal transmissions based on shifted frequency combs. These signals are optimized in terms of orthogonality. The period of the comb is related to the channel's time spreading in order to provide a simultaneous measure of the transfer functions between all sources and receivers. After explaining the construction of these signals and their properties, the effects of ocean fluctuations on detection using the D.O.R.T. method are illustrated through simulations. Finally a small-scale experiment validates the orthogonal family for detection in shallow water.

II. ADAPTIVE INSTANT RECORD SIGNALS

The most natural way to record different impulse responses between several points is to generate short signals successively from each emitting point and directly record all the impulse responses on the recording points. This technique is called time division multiplexing access (TDMA) and is a natural solution for speeding up the acquisition of all Green functions. In the case of multipath propagation, the time interval between two successive signals has to be greater than the dispersion time of the medium in order to avoid intersymbol interference. The impulse responses are then easily extracted from the received signals by time gates. This method is straightforward but provides a poor signal-to-noise ratio. Hence it would be better to send continuous codes simultaneously from all the emitting points and process the received signals by matched filter in order to extract the impulse responses.

While a ''universal'' family of codes is not realistic, a family of codes that is adapted to a specific application can indeed be developed. In this work, we propose a new family of codes considering primarily the physical meaning of the propagation in dispersive media. Indeed, the more dispersive a medium, the shorter the correlation length of the transfer function. All frequencies within one correlation length cross the same environment and carry the same information so that the measure of the transfer function at only one frequency within the correlation length is sufficient. Therefore, these codes are defined as a finite sum of sinusoidal functions with uniformly distributed frequencies. The different code spectra are switched in without crossover so to ensure orthogonality. The main idea of the technique is to choose the best spacing between the frequencies from knowledge of the time spreading of the propagation media. The inverse of the time spreading corresponds to the correlation length of the spectrum of the impulse response. In this way, if the spectra spacing is small compared with this correlation length, even if the different codes have access to different frequencies, all the impulse responses are well sampled in frequency and their reconstruction would be adapted to the propagation and the medium properties. In this theoretical part, we present an analytical description of these codes. We show quantitatively how to take into account the time spreading of the media. Later, we focus on the quality of the orthogonality with respect to the duration of the adaptive instant records signals (AIRS). It is shown that the signals have to be at least as long as the product between the time spreading and the number of AIRS of the family and are even better when the time spreading is overestimated. For the purpose of underwater acoustics, the signal-to-noise ratio is determined and analyzed. Finally we compare three particular AIRS based on pulse, random, and linear frequency modulation scheme. Finally, we compare the AIRS with Kasami signals, a classical set of orthogonal signals usually used for communication.

A. Impulse and nonoverlapping codes

The simplest situation where simultaneous signals can be used in order to record two different impulse responses consists of two emitting points 1 and 2 and one recording point. Thus two codes, $e_1(t)$ and $e_2(t)$, need to be determined and in order to be able to reconstruct the impulse responses between the two emitting points and the receiver, $e_1(t)$ and $e_2(t)$ should obey the system:

$$\begin{aligned} &e_1(t)\otimes e_1(-t)=\delta(t)\Leftrightarrow|e_1(\nu)|^2=1,\\ &e_2(t)\otimes e_2(-t)=\delta(t)\Leftrightarrow|e_2(\nu)|^2=1,\\ &e_1(t)\otimes e_2(-t)=0\Leftrightarrow e_1(\nu)e_2^*(\nu)=0, \end{aligned} \tag{1}$$

where $\otimes$ is the convolution operator $(f(t)\otimes g(t)=\int_{-\infty}^{\infty}f(\tau)g(t-\tau)d\tau)$. On the right is the Fourier transform of the time-relations written on the left. The first two relations allow extraction of the two different impulse responses and the third relation imposes no overlap between the two measurements. However, the third relation is not compatible with the first two. This incompatibility is strikingly evident on the Fourier transforms. Hence ''universal'' codes, which both have an impulsive auto-correlation function and covers different frequencies, do not exist.

B. Definition of AIRS

As explained in the introduction to the theoretical part, the different AIRS $e_k(t)$ are built on the discrete sum of sinusoidal functions. Then $e_k(t)$ is written:

$$e_k(t)=\mathrm{Re}\left[\sum_{l>0}e^{j(\phi_k^l-2\pi\nu_k^l t)}\right]\Pi_{[-T/2,T/2]}(t),\quad 0\leqslant k<N, \tag{2}$$

where Re means the real part, ϕ_k^l is any phase (random or not), and the frequency ν_k^l is defined as

$$\nu_k^l=l\,\delta\nu_1+k\,\delta\nu_2+\delta\nu_3, \tag{3}$$

where k, l are positive integers and $\delta\nu_1$, $\delta\nu_2$, and $\delta\nu_3$ are three constants. Practically speaking, in Eq. (2), the sum is performed over only a finite range limited by the bandwidth of the emitter/receiver devices. To ensure that there is no overlap between the signal spectra [Eq. (1)], it is assumed that $\delta\nu_1>N\delta\nu_2$.

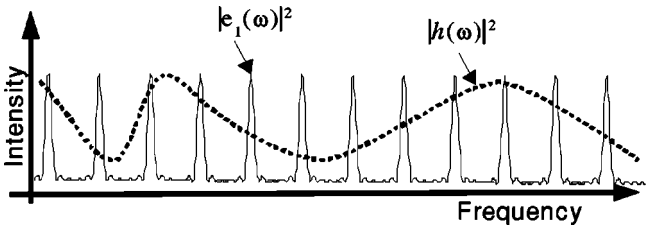


FIG. 1. Schematic view of AIRS in the frequency domain. The continuous curve corresponds to the Fourier transform of an impulse response.

The step function Π of Eq. (2) limits the codes to a finite length T. For simplicity, the codes are centered on time 0. It is instructive to compute the Fourier transform $(F(f)(\nu) = \int f(t)e^{2\pi j\nu t}dt, F^{-1}(f)(t) = \int f(\nu)e^{-2\pi j\nu t}d\nu)$ of this expression:

$$e_k(t) = T\,\mathrm{Re}\,F^{-1}\left[\sum_l e^{j\phi_k^l}\,\mathrm{sinc}((\nu-\nu_k^l)T)\right] = T\,\mathrm{Re}\,F^{-1}(\hat{E}_k(\nu)). \tag{4}$$

A schematic illustration of this equation is presented in Figs. 1 and 2 where all the quantities involved are outlined. The definition of $e_k(t)$ given in Eq. (4) means that the width of peak $\delta\nu_w$ is equal to $1/T$, which is the smallest possible width. One can also rewrite the AIRS as

$$e_k(t) \propto \mathrm{Re}\,F^{-1}\left[H(\nu)P(\nu)\xi\left(\frac{\nu-k\delta\nu_2}{\delta\nu_1}\right)\right]\Pi_{[-T/2,T/2]}(t), \tag{5}$$

where $\xi(\omega)$ is the Dirac Comb $(\xi(x) = \Sigma_{n=-\infty}^{+\infty}\delta(x-n))$, H is the Heaviside function $[H(\nu)=0$ when $\nu<0$ and $H(\nu)=1$ when $\nu>0]$, and p is a "function generator" with $|P(\omega)| = 1$ and $P(\omega_k^l) = e^{j\varphi_k^l}$.

Thanks to the convolution operator properties, it yields

$$e_k(t) \propto \mathrm{Re}\left(F^{-1}[H(\nu)P(\nu)] \otimes F^{-1}\left[\xi\left(\frac{\nu-k\nu_2}{\delta\nu_1}\right)\right]\right)\Pi_{[-T/2,T/2]}(t). \tag{6}$$

The inverse Fourier transform on the left-hand side of the convolution operator is by definition the Hilbert transform of the p function $(\mathrm{Hilbert}(p(t)) = \int_{\nu=0}^{\infty}\int_{\tau=-\infty}^{\infty}p(\tau)\mathrm{e}^{2\pi j\nu(\tau-t)}d\tau\,d\nu)$. The inverse Fourier transform on the right produces a Dirac comb of $1/\delta\nu_1$ period multiplied by a time-dependent phase factor due to the offset $\delta\nu_2$. Therefore $e_k(t)$ can be rewritten:

$$e_k(t) \propto \mathrm{Re}(\mathrm{Hilbert}(p(t)) \otimes [\xi(\delta\nu_1 t)e^{-2\pi kt\delta\nu_2}])\Pi_{[-T/2,T/2]}(t). \tag{7}$$

This equation gives another way to construct the kth orthogonal AIRS signal in the time domain: a generator function $p(t)$ of length $1/\delta\nu_1$ has to be chosen such as its Fourier transform obeys $|P(\omega)|=1$ and $P(\omega_k^l) = e^{j\varphi_k^l}$. Then one applies Hilbert transform on $p(t)$. The Hilbert transformed function is then replicated $N = T\delta\nu_1$ times. $1/\delta\nu_1$ time separates each replica. N rotating phase terms $\exp[2j\pi kl(\delta\nu_2)/\delta\nu]$ are applied to the N replicas, where l goes from 0 to $N-1$. Finally the real part of the result constitutes the $e_k(t)$ AIR signal. There are as many AIR signals as existing generator function, $p(t)$. Here, we focus on three useful generator functions, $p(t)$: a pulse, a linear frequency modulation (LFM), and a random function. The pulse is a good reference function in order to compare to two other generator functions. Later we show that the LFM function generates the most robust AIRS to the presence of external noise. The noise generator function would be very useful if one wants to acquire impulse responses discretely, the emitted signals looking like environmental noise.

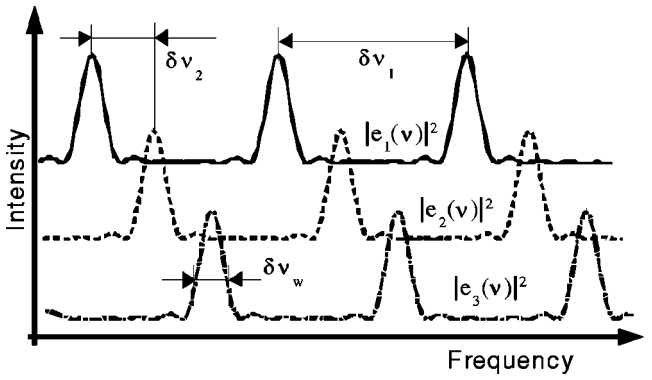


FIG. 2. Schematic view of the Fourier transform of three different AIRS: $e_1(t)$ (continuous), $e_2(t)$ dashed line, and $e_3(t)$ dot-dashed line.

C. Correlation between two AIRS

A matched filter performs the deconvolution task of the AIRS. The time equivalent of the matched filter is the correlation. The advantage of this deconvolution technique is its simplicity and the rapidity of the computations. The correlation between two AIRS is defined as follows:

$$C_{k,k'}(t) = e_k(t) \otimes e_{k'}(-t) \tag{8}$$

As the $\hat{E}_k(\nu)$ contain no negative frequencies, it can be shown from Eq. (4) that

$$C_{k,k'}(t) = \frac{T^2}{2}\mathrm{Re}\,F^{-1}[\hat{E}_k(\nu)\hat{E}_{k'}^*(\nu)], \tag{9}$$

which is written

$$C_{k,k'}(t) = \frac{T^2}{2}\mathrm{Re}\,F^{-1}\left[\sum_{l,l'} e^{j(\phi_l^k-\phi_{l'}^{k'})}\mathrm{sinc}(T(\nu-\nu_k^l)) \times \mathrm{sinc}(T(\nu-\nu_{k'}^{l'}))\right]. \tag{10}$$

1. Pulse compression: $k=k'$

Correct pulse compression is required in order to extract the impulse response from the matched filter. As shown in the following, we assume that when $k=k'$ overlapping between the different *sinc* functions is insignificant. The previous equation then becomes

$$C_{k,k}(t) = \frac{T^2}{2}\mathrm{Re}\,F^{-1}\left[\sum_l \mathrm{sinc}^2((\nu-\nu_k^l)T)\right]. \tag{11}$$

This can also be written

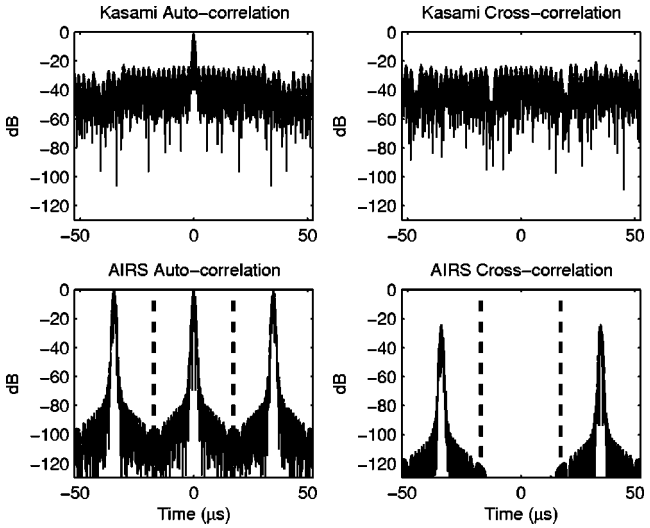


FIG. 3. Details of the cross (right) and auto (left) correlations using a 255-symbol Kasami sequence (upper subplots) and AIRS (lower subplots). The parameters to construct the two families are consistent with each other to allow comparison: in both cases the signal length is 560 μs, the number of elements is $N=16$. This gives a theoretical spreading time of $\tau=35$ μs used to build the AIRS. The vertical dashed lines indicate the $\pm\pi/2$ interval.

$$C_{k,k}(t)=\mathrm{Re}\left[\left(Te^{-j2\pi t(\delta\nu_3+k\delta\nu_2)}\Lambda\left(\frac{t}{T}\right)\xi(\delta\nu_1 t)\right)\otimes\left(\frac{\delta(t)}{2}-\frac{j}{2\pi t}\right)\right]. \tag{12}$$

The triangle function ($\Lambda(x)=\{1-|x|$ when $|x|<1$ and 0 elsewhere$\}$) $\Lambda(t)$ is simply the Fourier transform of the sinc^2. The Dirac comb, $\xi(\delta\nu_1 t)$, comes from the inverse Fourier transform of the discrete sum over l. Finally the right-hand side of the convolution operator is the inverse Fourier transform of the Heaviside function. Indeed, the sum is performed only over the positive frequency range in Eq. (11). Hence, as shown on the left of Fig. 3, working with a discrete spectrum implies that in place of a single Dirac function, the autocorrelation is a series of Dirac functions with a pitch equal to $1/\delta\nu_1$. If the impulse response of the medium is $h(t)$, the matched filter with the AIRS gives $C_{k,k}(t)\otimes h(t)$, so that after convolution, the signal is a succession of replica of the impulse response $h(t)$ separated by $1/\delta\nu_1$ with an amplitude factor. The replica of the $h(t)$ do not overlap if $\tau<1/\delta\nu_1$ where τ is the time spreading of $h(t)$. The central replica directly gives the impulse response because its amplitude factor equals 1. Strictly speaking, one should notice that, when considering limited bandwidth AIRS, the central Dirac of the autocorrelation is replaced by a peak of width given by the inverse of the bandwidth.

2. Code orthogonality

In order to avoid overlap between different AIRS, $C_{k,k'}(t)$ should be as close as possible to 0. In fact $C_{k,k'}(t)$ should be compared with $C_{k,k}(t)$, and to this end we compare the energy of both signals $E_{k,k'}/E_{k,k}$. We have

$$E_{k,k'}=\int|C_{k,k'}(t)|^2dt=\int|C_{k,k'}(\nu)|^2d\nu, \tag{13}$$

$$E_{k,k'}\approx\sum_{l,l',l''l'''}\int\mathrm{sinc}(T(\nu-\nu_k^l))\mathrm{sinc}(T(\nu-\nu_k^{l'}))\times\mathrm{sinc}(T(\nu-\nu_{k'}^{l''}))\mathrm{sinc}(T(\nu-\nu_{k'}^{l'''}))\times e^{j(\varphi_k^l-\varphi_k^{l'}+\varphi_{k'}^{l''}-\varphi_{k'}^{l'''})}d\nu. \tag{14}$$

Working out a good estimator of the crosstalk is quite difficult because it depends on the choice of the phases ϕ_k^l. However, the dominant part in this series of integrals corresponds to the terms where $l=l'=l''=l'''$, so that we can write in a good approximation

$$E_{k,k'}\approx\sum_l\int\mathrm{sinc}(T(\nu-\nu_k^l))^2\,\mathrm{sinc}(T(\nu-\nu_{k'}^l))^2d\nu. \tag{15}$$

To compare $E_{k,k'}$ and $E_{k,k}$ we compare each integral term in the sum. One term is equal to

$$I_{k,k'}=\int\mathrm{sinc}(T(\nu-(k-k')\delta\nu_2))^2\mathrm{sinc}(T\nu)^2d\nu,$$

while

$$I_{k,k}=\frac{2\pi}{3T}. \tag{16}$$

The highest term is obtained for $k'=k+1$. It is to notice that the orthogonally is greatly improved if tapering function is convolved to the generator function $p(t)$. Indeed, in such a case, frequency ripples of $P(\nu)$ are greatly decreased and the orthogonality is enhanced. The lower panel of Fig. 3 shows the autocorrelation $C_{k,k'}(t)$ between $\pm3\cdot\pi/2$ for $N=16$, $\tau=35$ μs, and $T=N\cdot\tau=560$ μs. The generator function $p(t)$ is a linear frequency modulation convolved with a Hanning window. The maxima of the cross-correlation $\max_{[-\pi/2,\pi/2]}\times\{C_{k,k'}(t)\}$ is approximately 120 dB below the autocorrelation pic $C_{k,k}(0)$.

3. Conditions for AIR codes

From the two preceding sections, to obtain N different and efficient AIRS means satisfying the following inequalities:

$$\frac{2}{T}\leqslant\delta\nu_2\leqslant\frac{\delta\nu_1}{N}\leqslant\frac{1}{N\tau}. \tag{17}$$

D. Robustness to external noise

In this section, we discuss the signal-to-noise ratio (SNR) that quantifies the robustness of the AIRS to the presence of external noise. For simplicity, we assume that the response for each channel is a temporal delta function $\delta(t)$: we assume that the codes are almost perfectly orthogonal. After simultaneously applying the signals $e_i(t)$ defined earlier on transducer number i, the recorded signal after propagation is $s(t)=\Sigma_i\beta_i e_i(t)+n(t)$, where $n(t)$ is an external white Gaussian noise and β_i an amplitude factor. The extraction of the kth impulse response yields to

TABLE I. Ratio for three different AIRS.

	Pulse	FML	Random
$\frac{\int e_k^2(t)dt}{e_{\max}^2}$	Pulse length	$\frac{T}{2}$	$<\frac{T}{2}$

$$s(t)\otimes e_k(-t)=\sum_{i=1}^{N}\beta_i e_i(t)\otimes e_k(-t)+n(t)\otimes e_k(-t), \tag{18}$$

$$s(t)\otimes e_k(-t)\approx\beta_{k'}e_k(t)\otimes e_k(-t)+n(t)\otimes e_k(-t). \tag{19}$$

At $t=0$, the intensity of the first term on the right-hand side of Eq. (18) is given by $(\beta_k\int e_k^2(t)dt)^2$. The intensity of the noise [second term in Eq. (18)] is $\langle n^2\rangle\tau_n\int e_k^2(t)dt$, where $\langle n^2\rangle$ is the pressure of the external noise variance and τ_n is the correlation time of the noise. Here τ_n is assumed to be very small. Finally, the SNR is equal to

$$\mathrm{SNR}=10\log\left(\frac{\beta_k^2\int e_k^2(t)dt}{\langle n^2\rangle\tau_n}\right). \tag{20}$$

Obtaining the best SNR means maximizing the β_k coefficient. The maximum pressure level, $P_{\max}$, that can reach the receiver is the limiting factor for β_k. It depends on the source power and the wave attenuation (dissipation, scattering, geometrical spreading, etc.). So the best β_k equals $P_{\max}/e_{\max}$ where $e_{\max}$ is the absolute maximum value of $e_k(t)$-AIRS. Thus the SNR becomes

$$\mathrm{SNR}=10\log\left(\frac{P_{\max}^2}{\langle n^2\rangle\tau_n}\cdot\frac{\int e_k^2(t)dt}{e_{\max}^2}\right) \tag{21}$$

and depends on the characteristic time

$$\frac{\int e_k^2(t)dt}{e_{\max}^2}$$

of the AIRS code. The three characteristic times corresponding to the three generator functions mentioned in Sec. II B are presented in Table I. Obviously, the longer the signal is, the larger

$$\frac{\int e_k^2(t)dt}{e_{\max}^2}$$

is, and the more robust to external noise are the codes. The most robust generator function is the frequency modulation.

E. Comparison between AIRS and PN sequence

The code division multiple access method has received considerable attention during the past several years. This method provides nearly orthogonal signals under several families such as Gold codes, m-sequences, and Kasami.[17] A particular AO/LSE initial-phase optimized long sequence of the Kasami set[18] has been chosen for comparison with the AIRS. In the ultrasonic domain, both signal lengths have been set to $T=560\,\mu$s with a common bandwidth of $B=1$ MHz at -3 dB around a center frequency of $f_c=1.5$ MHz. The Kasami sequence is of length $2^m-1=255\,(m=8)$ and the number of signals in each family is $N=2^{m/2}+1=16$. This gives an estimation of $\tau=35\,\mu$s used to build the AIRS. The cross and autocorrelation between two signals of the families are shown in Fig. 3. As expected, the AIRS minimize the cross-correlation down to approximately -120 dB in-between the central window duration of τ. On the other hand, the Kasami correlation shows a single peak through its entire duration, despite a much higher level of cross correlation at approximately[19]

$$-20\log_{10}\left(\frac{2^m-1}{N+1}\right)\approx-23\ \text{dB}.$$

This property allows the transmission of long communication sequences using the Kasami codes, while the AIRS are designed for precise measures of transfer functions of duration shorter than τ. Shallow water applications such as tomography or time reversal detection lead to the need for optimized signal to noise ratio, since the multipath structure usually shows strong contrast between direct rays and surface and bottom reflected rays.

F. Robustness to Doppler

AIRS are obviously sensitive to both frequency shift and frequency spreading found in shallow water propagation. AIRS are built using an estimation of the time spreading of the ocean. For shallow water propagation, the time spreading is closely related to the sea-state because of multiple surface reflections and surface noise generation. The spreading time decreases as the Doppler might increase with sea-state, allowing AIRS building with larger spacing between frequency peaks. The AIRS will have a limited overlapping due to Doppler spreading if the product of the spreading time and the number of codes satisfies

$$N\tau<\frac{1}{2\Delta f_{\text{Doppler}}}, \tag{22}$$

where $\Delta f_{\text{Doppler}}$ is the frequency spreading or frequency shift due to Doppler.

Roderick and Cron[20] have shown experimentally that under low and high sea-state conditions up to 7, the frequency spectrum consists of a carrier equal to the original transmitted frequency, with sideband frequencies related to the ocean spectrum and peaked at the frequency of the maximum energy of the surface waves. For all conditions of the ocean, this frequency spread is less than 1 Hz, and under low sea-state conditions, the spread is less than 0.2 Hz. In this last case of medium sea-states (3–4 Beaufort), the length of the codes $T=N\tau$ should be lower than 2.5 s, to compare with what one might expect considering the water depth, the mean wave height, and the ambient noise level.

Outside of major currents, for sources and receivers suspended from surface drifters or floats, typical velocities are 10 cm/s[21] creating 0.07 Hz Doppler shift at 1 kHz and average sound speed of 1500 m/s. The use of this kind of equipment limits $N\tau$ to approximately 7 s.

Although the use of lower frequencies would help, it is clear that AIRS must be designed carefully to compensate eventual Doppler effects from the instrumental setup or the environment.

III. APPLICATION TO THE D.O.R.T. METHOD

The acoustic configuration consists of a number of targets and one vertical source receiver array (SRA) of N transducers in a time invariant medium. Authors initially constructed the time reversal operator (TRO) by the direct and successive measurement of the interelement impulse functions $k_{i,j}(t)$ between transducers i and j and then Fourier transform these functions to provide the array response matrix $K(\omega)$.

The singular value decomposition of $K(\omega)$ gives a real diagonal matrix of singular values (square roots of the TRO eigenvalues) and eigenvectors. In this paper, we assume that the target behaves as isotropic scatterers. In this case, the number of ''nonzero'' eigenvalues is exactly the number of independent secondary sources given by the resolved scatterers present in the medium.[14] Eigenvectors are calculated in the time domain using an inverse Fourier transform, as described in Ref. 15.

A. Why are AIRS needed?

Time reversal is mainly based on the reciprocity of the acoustic pathways. For the same reason, the time reversal operator decomposition requires a stationary waveguide during both measurement of the interelement matrix and propagation of its eigenvectors. The ocean is far from a static environment. The problem arising from applying the D.O.R.T. method in an oceanic medium lies in the acquisition time. If M transducers are used for detection at range D, the acquisition of the interelement matrix lasts approximately

$$\Theta \approx \left(\frac{2D}{\overline{c_0}} + \tau_{\mathrm{rec}} \right) M \tag{23}$$

where $\overline{c_0}$ is the mean sound speed in the waveguide and τ_{rec} is the length of the reception window, at least as great as the sum of the length of the emitted signal and the time spreading of the waveguide. For example, if $M=51$ sets of signals, $\overline{c_0}=1500$ m/s at range $D=8$ km and the reception window length is around 1 s, Θ is longer than 10 min.

The question is: How sensitive is the D.O.R.T. method when the environment changes as quickly as fluctuations commonly observed, for example above the continental shelves in the northeast of the Atlantic, well known for strong tidal phenomenon? Tidal effects propagating through the continental shelf mainly govern summer oceanographic fluctuations in the sound speed profile. These fluctuations could be modeled by a so-called barotropic mode and baroclinic modes.[22] The barotropic mode models the sea-surface elevation due to external tides and induces a variation in the water depth. Added to this geometrical range independent deformation of the waveguide, the associated internal tides affect the distribution of the sound speed in the water column. The variation of the mean depth of the thermocline is due to the first baroclinic mode (Fig. 4) modeled as a plane wave propagating perpendicularly to the slope of the continental shelf.[22] The acoustic configuration is chosen therefore perpendicular to the propagation of the tides for range independent propagation in the presence of the barotropic and only the first baroclinic mode Ψ:

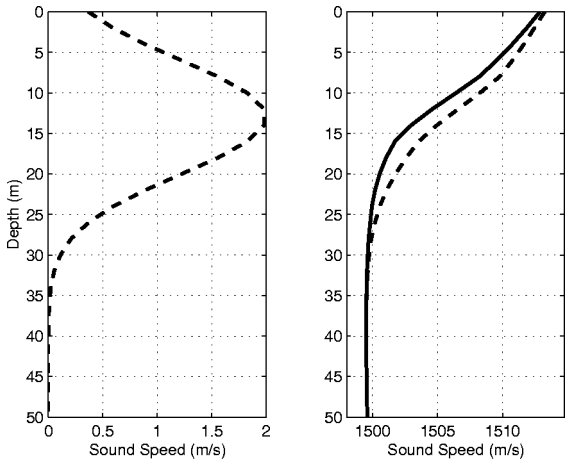


FIG. 4. First baroclinic mode $\Psi(z)$ used to model the internal tide (left). First (solid) and last (dotted) sound speed profiles used to evaluate the effects of environment variation during the acquisition of the time reversal operator (right).

$$\begin{aligned} &c(z,t) = c_0(z) + \alpha(t)\Psi(z), \\ &H(t) = H_0 + \alpha(t)\,\delta H, \\ &0 \leqslant \alpha(t) \leqslant 1, \end{aligned} \tag{24}$$

where $c(z,t)$ is the sound speed profile at depth z and time t, $H(t)$ the water depth at time t, $c_0(z)$ the mean sound speed profile (Fig. 4), δH the depth change of the water column during the interelement matrix acquisition, and α a real positive coefficient increasing linearly from zero to 1. A 51-element SRA is placed at 8 km range from two targets that are at depths of 5 and 20 m, respectively. The bottom is modeled as a single layer sandy bottom with compression speed 1750 m/s, absorption 0.7 dB/λ and density 1.9. The transfer matrix is measured while both the water depth goes up to $\delta H=5$ cm and the mean depth of the thermocline goes down to 1.98 m during the 10 min needed for the acquisition of a 51×51 interelement matrix. At each single source emission, the simulation is performed with Kraken[23] in a range independent configuration. As a reference, the $K(\omega)$ is also constructed and decomposed in a static environment where the sound speed profile is $c_0(z)$ and the water depth $H_0=50$ m.

Differences between both interelement matrices at 1 kHz due to the fluctuation of the environment are expected. For time variation, the major difference is in the loss of symmetry of the interelement matrix correlated with the loss of reciprocity between the sources and the receivers. The effect of the asymmetry on the decomposition is shown in Figs. 5 and 6 where additional noisy singular values appear and neither the focus of the first nor the second eigenvector occurs on the target positions anymore.

Detection capability in shallow water is clearly affected by the time varying properties and the time reversal operator

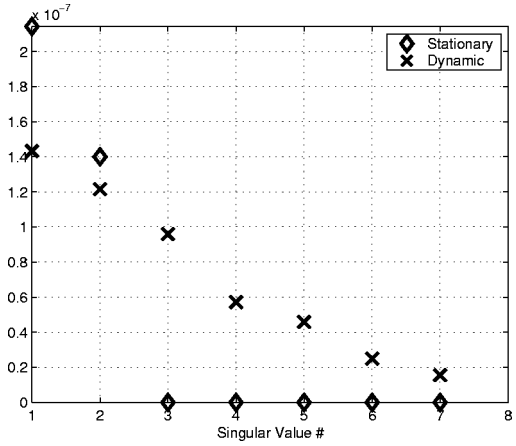


FIG. 5. First singular values obtained by decomposition of the time reversal operator in a stationary and varying environment.

decomposition method at sea needs a different kind of acquisition offered by adaptive instant record signals.

B. Experiment

The use of orthogonal signals is thus particularly suited for reducing the dependence of the method on fluctuations in the environment or movements of the instruments. The performance of these families of codes is evaluated experimentally for the detection of two targets in an ultrasound waveguide in the presence of bottom absorption and real reverberation, taking account of the transducer and electronic responses.

1. Experimental setup in an ultrasonic waveguide

The ultrasonic waveguide models in a realistic fashion oceanic propagation in shallow depths:[18] bottom absorption was produced by a sheet of Plexiglas™. The roughness of the bottom, the source of reverberation, was modeled by a layer of coarse sand a few millimeters thick, with grain size of 0.5 ± 0.2 mm, therefore of dimensions comparable to half

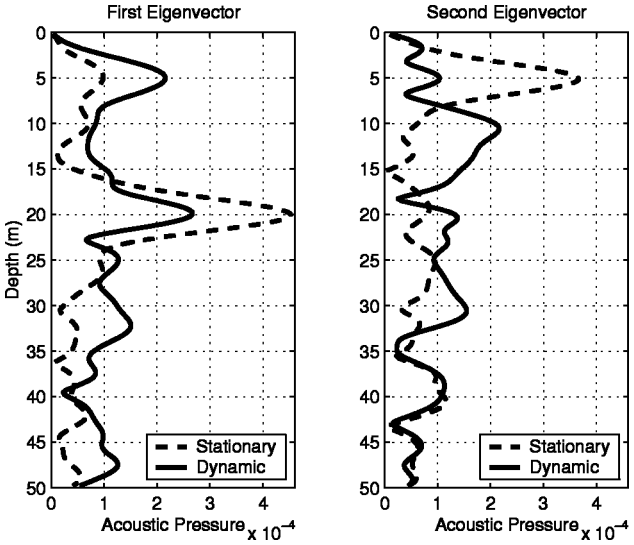


FIG. 6. Acoustic pressure vs depth obtained in the vertical target plane while propagating the first (left) and the second (right) eigenvector through a static (dotted) and varying (solid) waveguide.

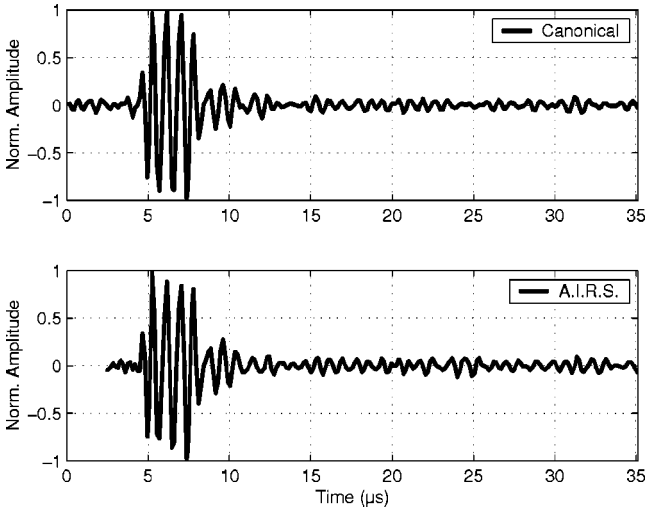


FIG. 7. Example of experimental Green function measured between transducers 11 and 12, directly (above) and synchronously using the AIRS (below).

the wavelength. The mean water depth was 30 mm. An array composed of $N=15$ transducers separated by 1.16 λ and having a central frequency of 1.5 MHz was placed vertically at one end of the guide. Two identical targets of diameter 0.8 λ were placed at two different depths at a range of 760 mm from the array. Random phased AIRS is chosen. This configuration (depth of water, propagation distance, critical angle of the bottom) allowed a value to be estimated for the guide's time spreading τ of 35 μs. Thus, the length of the AIR signals chosen was the shortest: $T=N\tau=525\,\mu$s.

2. Results

In order to assess the quality of the simultaneous measurement of $K^{\mathrm{AIRS}}(t)$, the interelement matrix, using AIRS, the interelement matrix $K^{\mathrm{can}}(t)$ was also measured in the canonical base. In the latter, the signal used was the first of the AIRS family used previously and a matched filter allowed comparable levels to be obtained.

Figure 7 shows the superposition of interelement responses between source number 11 and receiver number 12 measured either independently (canonical base) or at the same time as all other interelement responses (AIRS). The transfer functions obtained are very similar. The resemblance of the interelement responses was evaluated by calculating the N^2 intercorrelations between the functions:

$$\Xi_{k,k'}=\frac{\int|K^{\mathrm{can}}_{k,k'}(t)||K^{\mathrm{AIRS}}_{k,k'}(t)|dt}{\sqrt{(\int|K^{\mathrm{can}}_{k,k'}(t)|^2dt)(\int|K^{\mathrm{AIRS}}_{k,k'}(t)|^2dt)}}. \tag{25}$$

Figure 8 shows the distribution of the resemblance index $\Xi_{k,k'}$ between 0.73 and 0.97 around the mean value 0.89 synonymous with the experimental feasibility of the simultaneous acquisition of broadband interelement responses.

Both interelement matrices $K^{\mathrm{can}}(t)$ and $K^{\mathrm{AIRS}}(t)$ were decomposed between 0.75 and 2.25 MHz. The singular values obtained in the canonical base and using AIRS are shown in Fig. 9 as a function of frequency as a line and dotted line, respectively. The structure of the singular values remained

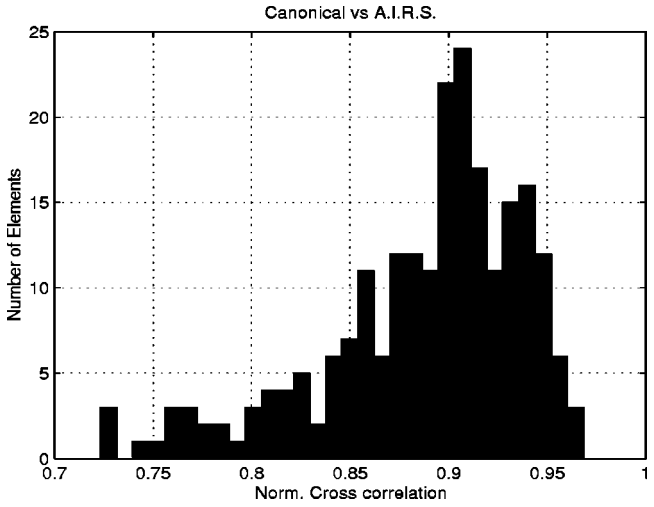


FIG. 8. Distribution of the 15×15 experimental resemblance index $\Xi_{k,k'}$.

unchanged despite the synchronous use of orthogonal signals. Two singular values clearly stand out from the others. Each of them corresponds to one of the two targets present in the water column. The third singular value is associated with the reverberation.[16] The singular values obtained in the two bases are similar. In particular, the levels of actual values associated with the reverberation remain independent of the acquisition base. The deviation between the canonical basis and AIRS for the first eigenvalue for frequencies above 1.5 MHz is likely due to weak nonlinearities of our electronics. Indeed using the canonical basis, the received signals are very weak and therefore distorded by our electronics. Nevertheless, reverberation plays a particular role in the synchronous acquisition of the time reversal operator. As the emitted signals are long, the level of near reverberation created by the end of the transmission may overlap and interfere significantly with the beginning of the target echoes. Even in the case of an intrinsically weak reverberation level, the difference in range covered by the target echoes and the near reverberation of the array may considerably upset the ability of the method to separate echoes from targets from those of reverberation. This robustness observed experimentally relative to bottom reverberation can certainly be explained by the averaging effect due to the spatial incoherence of the reverberation when compared with the spatial coherence of the target echoes.

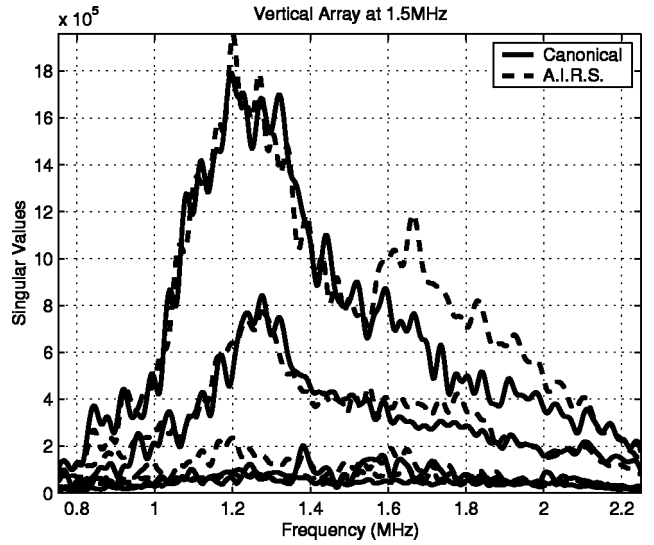


FIG. 9. Singular values obtained by decomposition of TRO acquired both in the canonical base (continuous line) and simultaneously while the 15 transducers are emitting together.

IV. CONCLUSION

Time reversal techniques are an example of the increasing importance of obtaining a large number of transfer function measurements between transducers. To satisfy the stationarity conditions of the propagation medium and limit the effects of displacement of the instruments, these acquisitions must be made in the shortest possible period of time. To comply with this constraint while optimizing the signal-to-noise ratio, orthogonal signals can be constructed from an estimation, even a rough one, of the medium's time spreading. Transmitted simultaneously by all the sources, exact broadband transfer functions are obtained by successive correlations and windowing of the signals measured by the hydrophones. These adaptive instant record signals transmitted simultaneously by all the transducers are, by construction, robust to guided propagation. Contrary to classical orthogonal signals such as Kasami or m-sequences used in communication, AIR signals gives sufficient othogonality (as shown down to -120 dB) to provide both measure of the direct arrivals and the multiple reflected arrivals within the spreading timeframe. Offering this optimized orthogonality only in this short timeframe clearly prohibits any use for underwater communication applications. Nevertheless, this property does not limit in any case the measure of transfer functions always obtained in the central window of the cross correlation even if the source arrays and receiver array are placed at different locations in the waveguide. In fact, the only need is that the transfer function issue from the desire sources on the desire receivers has a delay spread that falls within a window of duration τ. In practice, the results of the experiment even demonstrate the match between the multipath structure obtained at long ranges in the shallow water waveguide both with a direct measure and the backscattering of the AIRS. The multi-target detection provided experimentally by the decomposition of the time reversal operator gives the same resolution using the AIRS even if the performance given by this method is based on the quality of the measure of the multipath structure.

However, as the experimental results demonstrate, the presence of bottom absorption and reverberation is not an obstacle to the performance of these orthogonal signals, which should offer a new perspective to multiple input multiple output tomography applications.

[1] D. R. Jackson and D. R. Dowling, "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**, 171–181 (1991).
[2] D. R. Dowling, "Phase-conjugate array focusing in a moving medium," J. Acoust. Soc. Am. **94**, 1716–1718 (1993).
[3] D. R. Dowling and D. R. Jackson, "Coherence of acoustic scattering from a dynamic rough-surface," J. Acoust. Soc. Am. **93**, 3149–3157 (1993).
[4] S. R. Khosla and D. R. Dowling, "Time-reversing array retrofocusing in

simple dynamic underwater environments,'' J. Acoust. Soc. Am. **104**, 3339–3350 (1998).
[5] M. R. Dungan and D. R. Dowling, ''Computed narrow-band time-reversing array retrofocusing in a dynamic shallow ocean,'' J. Acoust. Soc. Am. **107**, 3101–3112 (2000).
[6] W. A. Kuperman, W. S. Hodgkiss, H. C. Song *et al.*, ''Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror,'' J. Acoust. Soc. Am. **103**, 25–40 (1998).
[7] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, ''A long-range and variable focus phase-conjugation experiment in shallow water,'' J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[8] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, ''Iterative time reversal in the ocean,'' J. Acoust. Soc. Am. **105**, 3176–3184 (1999).
[9] J. S. Kim, H. C. Song, and W. A. Kuperman, ''Adaptive time reversal mirror,'' J. Acoust. Soc. Am. **109**, 1817–1825 (2001).
[10] S. Kim, G. F. Edelmann, and W. A. Kuperman, ''Spatial resolution of time-reversal arrays in shallow water,'' J. Acoust. Soc. Am. **110**, 820–829 (2001).
[11] S. Kim, W. A. Kuperman, and W. S. Hodgkiss, ''Robust time reversal focusing in the ocean,'' J. Acoust. Soc. Am. **114**, 145–157 (2003).
[12] G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, ''An initial demonstration of underwater acoustic communication using time reversal,'' IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[13] C. Prada and M. Fink, ''Eigenmodes of the time reversal operator: A solution to selective focusing in multiple-target media,'' Wave Motion **20**, 151–163 (1994).
[14] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, ''Decomposition of the time reversal operator: Application to detection and selective focusing on two scatterers,'' J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[15] N. Mordant, C. Prada, and M. Fink, ''Highly resolved detection and selective focusing in a waveguide using the D.O.R.T. method,'' J. Acoust. Soc. Am. **105**, 2634–2642 (1999).
[16] T. Folégot, C. Prada, and M. Fink, ''Resolution enhancement and separation of reverberation from target echo with the time reversal operator decomposition,'' J. Acoust. Soc. Am. **113**, 3155–5160 (2003).
[17] D. V. Sarwate and M. B. Pursley, ''Cross correlation properties of pseudorandom and related sequences,'' Proc. IEEE **68**, 593–619 (1980).
[18] K. H. Karkkainen, ''The influence of initial-phases of a PN code set on the performances of an asynchronous DS-CDMA system,'' Kluwer Academic Publishers in Wireless Personal Communications Vol. 13 (Kluwer Academic, Dordrecht, 2000), pp. 279–293.
[19] J. G. Proakis, *Digital Communications, Mc-Graw Hill Series in Electrical and Computer Engineering* (McGraw-Hill, New York, 2001), p. 905.
[20] W. I. Roderick and B. F. Cron, ''Frequency spectra of forward-scattered sound from the ocean surface,'' J. Acoust. Soc. Am. **48**, 759–766 (1970).
[21] W. Munk, P. Worcester, and C. Wunsh, *Ocean Acoustic Tomography* (Cambridge University Press, New York, 1995).
[22] Y. Stéphan, X. Demoulin, T. Folégot, S. M. Jesus, M. B. Porter, and E. Coelho, ''Interaction of sound and tides,'' in *Experimental Acoustic Inversion Methods* (Kluwer Academic, Dordrecht, 1999), pp. 19–38.
[23] M. B. Porter, ''The KRAKEN normal mode program,'' Naval Research Laboratory Report No. NRL-9-6920, 1992.

The effect of stimulus-frequency ratio on distortion product otoacoustic emission components

Sumitrajit Dhar[a)]
Hugh Knowles Center, Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208

Glenis R. Long
Graduate Center, City University of New York, New York, New York 10016

Carrick L. Talmadge
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

Arnold Tubis
Institute for Nonlinear Science, University of California—San Diego, La Jolla, California 92093

(Received 25 January 2005; revised 11 March 2005; accepted 11 March 2005)

A detailed measurement of distortion product otoacoustic emission (DPOAE) fine structure was used to extract estimates of the two major components believed to contribute to the overall DPOAE level in the ear canal. A fixed-ratio paradigm was used to record DPOAE fine structure from three normal-hearing ears over a range of 400 Hz for 12 different stimulus-frequency ratios between 1.053 and 1.36 and stimulus levels between 45 and 75 dB SPL. Inverse Fourier transforms of the amplitude and phase data were filtered to extract the early component from the generator region of maximum stimulus overlap and the later component reflected from the characteristic frequency region of the DPOAE. After filtering, the data were returned to the frequency domain to evaluate the impact of the stimulus-frequency ratio and stimulus level on the relative levels of the components. Although there were significant differences between data from different ears some consistent patterns could be detected. The component from the overlap region of the stimulus tones exhibits a bandpass shape, with the maximum occurring at a ratio of 1.2. The mean data from the DPOAE characteristic frequency region also exhibits a bandpass shape but is less sharply tuned and exhibits greater variety across ears and stimulus levels. The component from the DPOAE characteristic frequency region is dominant at ratios narrower than approximately 1.1 (the transition varies between ears). The relative levels of the two components are highly variable at ratios greater than 1.3 and highly dependent on the stimulus level. The reflection component is larger at all ratios at the lowest stimulus level tested (45/45 dB SPL). We discuss the factors shaping DPOAE-component behavior and some cursory implications for the choice of stimulus parameters to be used in clinical protocols. [DOI: 10.1121/1.1903846]

PACS numbers: 43.64.Bt, 43.64.Ha, 43.64.Jb [BLM] Pages: 3766–3776

I. INTRODUCTION

Kim's (1980) initial prediction of two sources for the distortion product otoacoustic emission (DPOAE) at the frequency $2f_1-f_2$ (f_1 and f_2 are the frequencies of the stimulus tones and $f_1<f_2$) has undergone considerable development and formalization over the last decade. The DPOAE at $2f_1-f_2$, as well as other (apical) DPOAEs where $f_{dp}<f_1<f_2$, are understood to be a vector sum of (at least) two components, each generated at disparate locations on the basilar membrane due to distinct mechanisms. The nonlinear interaction between the stimulus tones around the tonotopic region of the higher-frequency stimulus tone (f_2) generates the *distortion* or *generation* component. The *reflection* component, on the other hand, is due to linear coherent reflections from a randomly distributed roughness about the tonotopic region of the DPOAE. Detailed analytic descriptions of the above model, or parts thereof, have been presented by several groups (e.g., Zweig and Shera, 1995; Talmadge *et al.*, 1998; Mauermann *et al.*, 1999) along with several lines of supporting experiments. Although these models apply to all apical DPOAEs, we will use the term DPOAE synonymously with the DPOAE at the frequency $2f_1-f_2$ in this paper. Evidence in support of such a two-source, two-mechanism model has come from suppression experiments (e.g., Kummer *et al.*, 1995; Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Kalluri and Shera, 2001), results in the pseudotime domain following Fourier analysis of frequency-domain data (e.g., Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001, 2002; Knight and Kemp, 2001; Dhar *et al.*, 2002), as well as true time-domain recordings of DPOAEs (Talmadge *et al.*, 1999). An important consequence of the interaction of energy from two sources is the DPOAE fine structure (a pseudoperiodic variation of the DPOAE level and phase as a function of frequency).

In a view parallel to that of the two-source, two-mechanism dichotomy, the two DPOAE components have

[a)]Corresponding author: Sumitrajit Dhar. Electronic mail: s-dhar@northwestern.edu

also been referred to as *place* and *wave* fixed. Initially proposed as two sources of DPOAEs originating at the same location on the basilar membrane (Kemp, 1986), the place-/wave-fixed description has embraced the mechanistic and location differences between DPOAE components; with the wave- and place-fixed components equated to the generator and reflection components, respectively (Knight and Kemp, 2000, 2001).

A third view has also recently emerged, driven primarily by experimental comparisons of delay measures of otoacoustic emissions and basilar membrane recordings (Ruggero, 2004; Ren, 2004). Here, otoacoustic emissions (OAEs) are proposed to (i) be generated from a distributed area extending toward the base of the cochlea similar to that associated with cochlear microphonics; and (ii) propagate via acoustic compression waves in the cochlear fluids rather than as traditional traveling waves on the basilar membrane. Arguments for this view come from the presence of OAEs in basilar-membraneless ears of frogs and limited sets of data from gerbils, guinea pigs, and chinchillas (Ruggero, 2004). While these are interesting developments, regularly observed features of OAEs such as phase behavior and fine structure are yet to be explained using this new model. The significance of Ren's (2004) report is also unclear, given his choice of stimulus-frequency ratios for which the propagation of DPOAE energy from the f_2 CF region is known to be predominantly in the apical direction (Schneider *et al.*, 2003). Moreover, evidence in support of this model is yet to be presented from human ears. Our work here, representing experimental results from normal-hearing human ears, is presented using the theoretical framework of the traditional ''two-source'' model of apical DPOAEs.

Talmadge *et al.* (1998) provide a detailed derivation of a ''two-source'' model of the DPOAE recorded in the ear canal (P_e) at the angular frequency (ω_{dp}). Two components of the distortion product (DP) are initially generated in the cochlear region of maximum overlap of the waves from the two stimulus tones. One component (P_l) travels basally (backward) toward the cochlear base and into the ear canal where it gives a DPOAE contribution $P_l(\omega_{\mathrm{dp}})$. The other generated DP component in the cochlea travels apically (forward) to the DP characteristic frequency (CF) region, where it is partially reflected toward the cochlear base into the ear canal via coherent scattering from assumed cochlear inhomogeneities (Zweig and Shera, 1995; Talmadge *et al.*, 1998). This reflected energy results in an ear-canal signal contribution of $P_r(\omega_{\mathrm{dp}})R_a(\omega_{\mathrm{dp}})$, where $P_r(\omega_{\mathrm{dp}})$ is related to the initial apical moving DP component in the cochlea and R_a is the apical reflectance. In a simplified version of the model it is assumed that all of the reflected energy leaves the cochlea and is not reflected back into the cochlea from the junction of the cochlea and the middle ear. Consequently the DPOAE pressure signal recorded in the ear canal P_e at the frequency ω_{dp} can be represented as

$$P_e(\omega_{\mathrm{dp}}) = P_l(\omega_{\mathrm{dp}}) + R_a(\omega_{\mathrm{dp}})P_r(\omega_{\mathrm{dp}}). \quad (1)$$

Later versions of the model (supported by experiments) include multiple internal reflections between the stapes and the DPOAE CF region (Dhar *et al.*, 2002). However, for the purposes of this paper we assume that the reflectance at the oval widow (R_b in our models) equals zero. We will refer to the representation of P_l in the ear canal as the *generator* component and that of R_aP_r as the *reflection* component.

In this paper we evaluate the relative contributions of the generator and reflection components to the ear-canal DPOAE. Specifically we wish to examine the behavior of these two components as a function of stimulus-frequency ratio. The most comprehensive work in this area has been presented in a series of three articles by Knight and Kemp (1999, 2000, 2001). In the first of the series, they demonstrated similarity between the characteristics of transient evoked OAEs (TEOAEs) and DPOAEs generated for small (*narrow*) stimulus-frequency ratios, alluding to concordance between their generation processes. In the second publication, a detailed analysis of the phase behaviors of multiple orders of DPOAEs was used to demonstrate that the mechanism for DPOAE generation was dominantly place- and wave-fixed for narrow and wide stimulus-frequency ratios, respectively. In the third publication, these authors confirmed these results using the same data transformed into the time domain using inverse Fourier transformation. The essence of the Knight and Kemp findings can be summarized as (i) DPOAE generation is predominantly place fixed at narrow ratios; (ii) the wave- and place-fixed components demonstrate a bandpass shape as a function of stimulus-frequency ratio; (iii) the bandpass shape of the place-fixed component is asymmetrical, the ''drop-off'' being much steeper at the wide ratios; and (iv) the change in dominance of one mechanism over the other occurs approximately around a stimulus-frequency ratio of 1.1. Note that we refer to the place- and wave-fixed components as the reflection and generator components, respectively.

Schneider *et al.* (2003) investigated the dominance of the components in the guinea pig by examining the DPOAE group delay. In contrast to the work reported by Knight and Kemp, where DPOAE recordings were done with fixed stimulus-frequency ratios, Schneider *et al.* varied the frequency of one stimulus tone while keeping the other fixed in frequency. The transition in dominance from place- to wave-fixed mechanisms occurred at a stimulus-frequency ratio wider than that observed in human ears, consistent with the difference in the sharpness of basilar-membrane tuning between guinea pigs and humans as well as wider fine-structure spacing in rodents (Long *et al.*, 1999; Withnell *et al.*, 2003). Schneider *et al.* (2003) demonstrate the relationship between the two components at stimulus-frequency ratios greater than 1.35 to be ambiguous.

In sharp contrast to the somewhat extensive experimental work presented in this area to date, a theoretical examination of the mechanistic implications of these findings has been limited. Knight and Kemp (1999) base the explanation of their results on the rapid phase variation of the $2f_1-f_2$ wave near the f_2 peak region as f_2/f_1 approaches 1. This variation, the authors point out, will cause a reduction of the magnitude of the backward traveling wave as DPOAE wavelets will fail to combine in phase. While these phase-related arguments may be valid, amplitude effects as a function of changing stimulus-frequency ratio cannot be neglected.

Shera (2003) has included amplitude effects in his treatment of the subject and provided a more complete analytic description of the interplay between the generator and reflection components as a function of the changing stimulus-frequency ratio. These issues, in light of our data, are presented in detail in Sec. IV.

In this paper we extend previous findings by evaluating DPOAE component behavior as a function of stimulus-frequency ratio in normal-hearing human ears. We report novel behavior of DPOAE components at wide stimulus frequency ratios. Our study also differs from previous reports in the use of lower level stimulus tones, which reveals features of DPOAE components not reported previously. We will conclude by providing a theoretical analysis of our results and preliminarily comment on the implications of these findings on the clinical application of DPOAEs.

II. METHODS

A. General methods

The methods used to record and analyze data were similar to those reported by Dhar *et al.* (2002). In brief, high-resolution DPOAE data were recorded from three normal-hearing subjects recruited from the Purdue University community. These three subjects were selected from a total of approximately ten subjects screened. Subject selection was based on a negative family history of hearing loss, normal middle ear function as determined by type-A tympanograms, hearing thresholds better than 10 dB HL at half-octave frequencies measured using tracking audiometry between 250 and 8000 Hz, and a ''desirable'' OAE profile. OAE screening involved recording of spontaneous OAEs (SOAEs) as well as DPOAEs recorded using stimulus levels of 65 dB SPL and a stimulus-frequency ratio (f_2/f_1) of 1.225; DPOAE results were considered ''desirable'' when the signal-to-noise ratio between DPOAE-level maxima and the noise floor was approximately 20 dB or greater and distinct fine structure was observed in some frequency ranges. The presence or absence of SOAEs was not used to determine eligibility, but the frequency of the SOAE(s) when present was considered during an examination of the DPOAE fine structure. Subjects were seated comfortably in a recliner inside a double-walled IAC sound chamber for the experimental sessions, that lasted approximately 2 h each.

All experiments were conducted using custom designed software developed in the laboratory running on a NeXT workstation. Signals generated on the computer were passed through a set of TDT PA4 programmable attenuators and a HB6 headphone buffer before being fed to a set of ER2 tube phones. The signals were delivered to the subjects ear canal from the tube phones through a ER10B microphone probe assembly. Signals recorded in the ear canal by the ER10B microphone were passed through the microphone's preamplifier and a battery-operated Stanford Research SR560 low-noise amplifier/filter. The output of the SR560 was digitized using a Singular Solutions AD64x at a sampling rate of 44 100 Hz before being stored on the hard disk of the NeXT computer for offline analysis. The ER10B probe was fitted to the subject's ear using a GSI emittance probe tip matched in size to the ear canal. Stimulus levels were calibrated in a Zwislocki coupler and system distortion, measured in an aluminum 2 cm^3 coupler, was generally below -30 dB SPL in the frequency range of interest for the highest stimulus levels used.

Extensive DPOAE recordings were obtained at various stimulus levels and frequency ratios from three subjects who satisfied the above mentioned inclusion criteria. The following stimulus frequency ratios were used: 1.053, 1.065, 1.08, 1.11, 1.14, 1.18, 1.22, 1.26, 1.30, 1.32, 1.34, and 1.36. Three equal-level stimulus pairs ($L_1=L_2=45$, 65, 75 dB SPL) were used to record data. Additionally, several L_2 (45, 50, 55, and 60 dB SPL) were used in conjunction with an L_1 of 65 dB SPL. The choice of stimulus levels was made to cover a range of input levels. A comprehensive range of stimulus-frequency ratios were chosen consistent with our principal aim—examination of DPOAE-component characteristics as a function of stimulus-frequency ratio. The choice of stimulus levels was less rigorous, and was made simply to provide a range of unequal and equal level inputs. A resolution of 0.025 mm based on the Greenwood map (Greenwood, 1990) was used for frequency spacing. This resulted in approximate frequency resolutions of 4 Hz around 1500 Hz and 8 Hz around 2500 Hz. Each fine structure period was sampled by 16 data points. The fixed-ratio data sets were converted from the frequency domain to the time domain using a custom designed IFFT algorithm and the DPOAE components were isolated using time windowing (Stover *et al.*, 1996; Knight and Kemp, 2001; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001). The reader is directed to Dhar *et al.* (2002) for a complete description of the IFFT algorithm and the time-windowing protocol. An example of the transformations used in estimating DPOAE-component magnitudes for subject AE is displayed in Fig. 1.

Frequency domain amplitude and phase data (Fig. 1; panel A) were processed through an IFFT protocol to yield a time-domain representation (Fig. 1; panel B). These data were time windowed to isolate the generator and reflection components. Each component was then converted back to the frequency domain (Fig. 1; panel C). Note that the limits of our time-domain filters were chosen to isolate the principal DPOAE components only. Specifically, the upper limit of the second window was carefully selected, thereby attempting to eliminate any ''contamination'' of the reflection component by subsequent intracochlear reflections. In practice, subsequent reflections, when present, were analyzed using independent filters; those data, irrelevant to the current issue, are not presented here. The magnitudes of the components, obtained from the reconstructed frequency spectra, were sampled every 24 Hz to yield an average value for each stimulus condition.

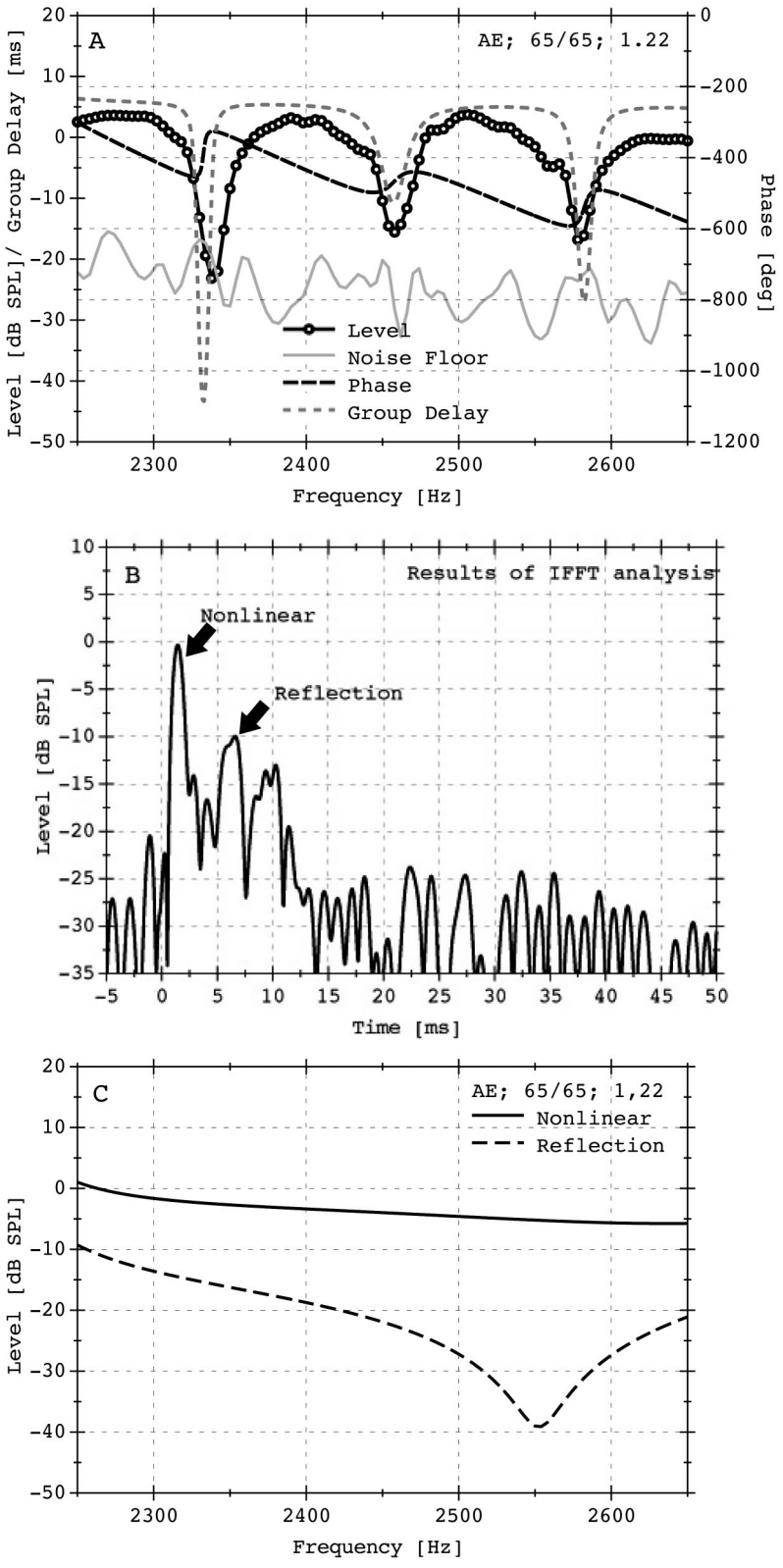


FIG. 1. An example of the process followed to determine magnitudes of DPOAEs components. DPOAE amplitude, phase, group delay, and noise floor are displayed in panel A for subject AE ($L_1 = L_2 = 65$ dB SPL; $f_2/f_1 = 1.22$). These data were processed through an IFFT algorithm to transform them to the time domain. The results of the IFFT analysis for these data are presented in panel B. The components were isolated in time and converted back to the frequency domain (panel C). DPOAE component magnitudes were estimated from these frequency-domain data.

B. Statistical Methods

A repeated measures multifactor analysis of variance (MANOVA) was performed to evaluate the effect of stimulus parameters on component magnitudes. While the magnitudes of the individual components served as dependent variables, stimulus level and frequency ratio were the independent variables. Frequency was used as an observational variable as we recorded data over different frequency ranges for the three subjects to maximize the observation of fine structure, and hence was not included in the model. Separate analyses were performed for the generator and reflection components to avoid interfering with the orthogonality of the models as

independence of the two components could not be easily established. DPOAE-component magnitudes were studied between 2250 and 2650 Hz for subject AE, between 1550 and 1950 Hz for subject CZ, and between 1800 and 2200 Hz for subject KT.

III. RESULTS

A. General results

The main effects of subject ($F=125.9$, $df=2$), stimulus level ($F=483.89$, $df=6$), stimulus-frequency ratio ($F=198.27$, $df=11$) and the interaction of level and ratio ($F=13.19$, $df=63$) on the generator component (A_{gen}) were all highly significant ($p<0.0001$). A similar analysis for the reflection component (A_{ref}) also showed highly significant ($p<0.0001$) main effects of subject ($F=597.99$, $df=2$), stimulus level ($F=171.68$, $df=6$), stimulus-frequency ratio ($F=109.16$, $df=11$), and the interaction between level and ratio ($F=9.34$, $df=63$). Note that data recorded with all stimulus conditions with the exception of $L_1=L_2=45$ dB SPL, $f_2/f_1>1.3$ were included in these statistical analyses. Each subject showed differences in the relative amplitudes of the two components across frequency. Interactions between the variable subject and other independent variables were not included in the model as different subjects contributed data from different frequency ranges. Thus, the effects of frequency and subject could not be isolated in this model.

B. Effects of stimulus frequency ratio

The results of a Tukey Studentized Range (HSD) (Inc., 1999) test comparing mean values of A_{gen} and A_{ref} for different stimulus-frequency ratios are presented in Table I. The top and bottom halves of the table present comparisons for A_{gen} and A_{ref}, respectively. While the last two columns of Table I present the computed means for each ratio and rank them in descending order, the initial columns show their Tukey letter groupings. Means sharing the same letter grouping were not significantly different ($\alpha=0.05$). Thus, the mean values of A_{gen} for stimulus ratios 1.22, 1.18, and 1.26 *did* differ significantly from each other as they have different letter groupings (A, B, and C in this case). However, A_{gen} means for 1.14, 1.32, and 1.11 share the same letter grouping (D) and hence *do not* differ significantly. This is suggestive of a fairly symmetrical bandpass shape. A_{gen} was highest for a ratio of 1.22, while A_{ref} was highest for 1.11. Additionally, the ratios generating the four highest values of A_{ref} *did not* differ significantly from each other, consistent with an asymmetric bandpass shape and/or greater variability between ears and across frequency within a given ear. The lowest values of A_{gen} and A_{ref} were generated at narrow and wide ratios, respectively. The strong statistical overlap between ratios generating high A_{ref} contrasted with the significant differences between 1.22, 1.18, and 1.26 for A_{gen} provides an indication of distinct differences between the magnitude functions for the two components.

Mean values for A_{gen} and A_{ref} for different ratios, collapsed across subjects, frequency, and stumulus levels, are presented in Fig. 2. The solid lines represent A_{gen}, while the dashed lines represent A_{ref}. The error bars represent ± 1 standard error. The lighter lines represent means calculated from the entire dataset, while the darker lines represent means calculated from the dataset after excluding data for stimulus levels of $L_1=L_2=45$ dB SPL. Data could not be recorded at all ratios for stimulus levels of 45/45 dB SPL, which resulted in a disparity in cell sizes for different ratios. Consequently, data acquired for stimulus levels of $L_1=L_2=45$ dB SPL were excluded from subsequent plots where data were collapsed across stimulus levels. Data with and without DPOAEs recorded with stimulus levels of $L_1=L_2=45$ dB SPL are presented in Fig. 2 to enable a comparison.

An examination of Fig. 2 reveals that the exclusion of data for stimulus levels of 45 dB SPL causes an upward shift of the traces at the ratios where we were able to record data at 45 dB, without significantly altering the slopes or the peak ratios for either A_{gen} or A_{ref}. The inclusion of DPOAEs recorded with stimulus levels of 45 dB SPL affected the curve for A_{gen} to a greater degree. The A_{gen} ratio function appears narrower and more peaked than that of A_{ref}. A_{ref} is higher than A_{gen} at narrow ratios and the two functions cross between the ratios of 1.11 and 1.14. The A_{gen} ratio function peaks around a ratio of 1.22. The two functions approach each other at ratios of greater than 1.3.

A_{gen} and A_{ref} for individual subjects are displayed in Fig. 3. In keeping with the general trend seen in Fig. 2, A_{ref} is

TABLE I. Results of the Tukey Studentized Range (HSD) test for generator (A_{gen}) and reflection (A_{ref}) components for the independent variable ratio. The results are in accordance with the following parameters: $\alpha=0.05$, $df=2724$, MSE=33.234 91 (A_{gen}) and 36.228 47 (A_{ref}), Minimum significant difference=1.7495 (A_{gen}) and 1.8266 (A_{ref}). Ratios at the extema *(1.053, 1.36)* and midpoint (**1.22**) of the distribution are specially marked; note the different positions of these ratios in the rank hierarchy.

Results for generator component A_{gen}

				Tukey grouping			Mean	Ratio
						A	−2.8184	**1.22**
					B		−4.6063	1.18
				C			−7.4451	1.26
			D				−10.0574	1.14
		E	D				−10.3632	1.32
		E	D				−11.7540	1.11
		E					−11.8192	1.30
	F						−14.3388	1.34
G	F						−16.0069	1.08
G							−16.1129	*1.36*
H							−19.0315	1.065
H							−19.3439	*1.053*

Results for reflection component A_{ref}

				Tukey grouping			Mean	Ratio
						A	−8.6516	1.11
						A	−8.8938	1.18
					B	A	−9.9972	1.14
					B	A	−10.1653	**1.22**
				C	B		−11.3941	1.08
				C			−12.0360	1.065
				C			−12.2809	*1.053*
			D				−14.5252	1.32
			D				−15.7227	1.26
		E	D				−16.2043	1.30
		E					−17.8815	1.34
	F						−20.8783	*1.36*

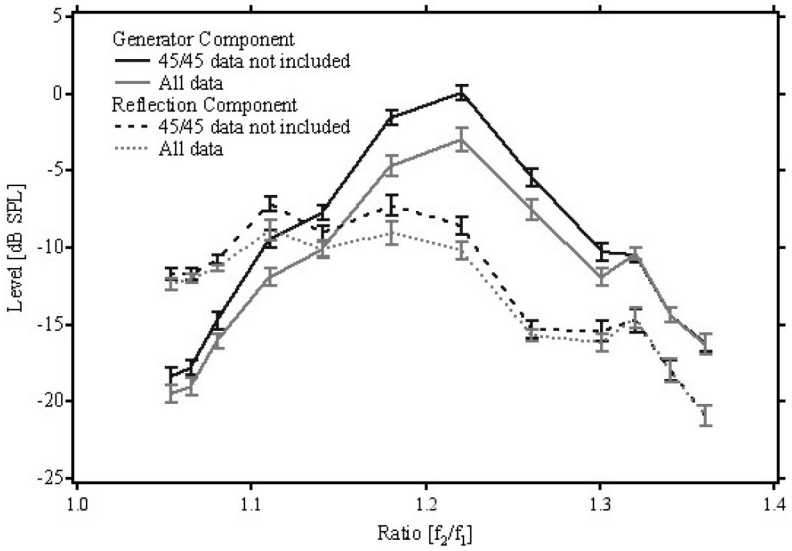


FIG. 2. Magnitude of DPOAE components as a function of the changing stimulus-frequency ratio. The solid and dashed lines indicate the magnitude of the generator (A_{gen}) and reflection (A_{ref}) components, respectively. The lighter lines represent means calculated on the entire dataset. The darker lines represent means calculated after excluding data points recorded with stimulus levels of $L_1 = L_2 = 45$ dB SPL. The error bars on all traces represent ± 1 standard error.

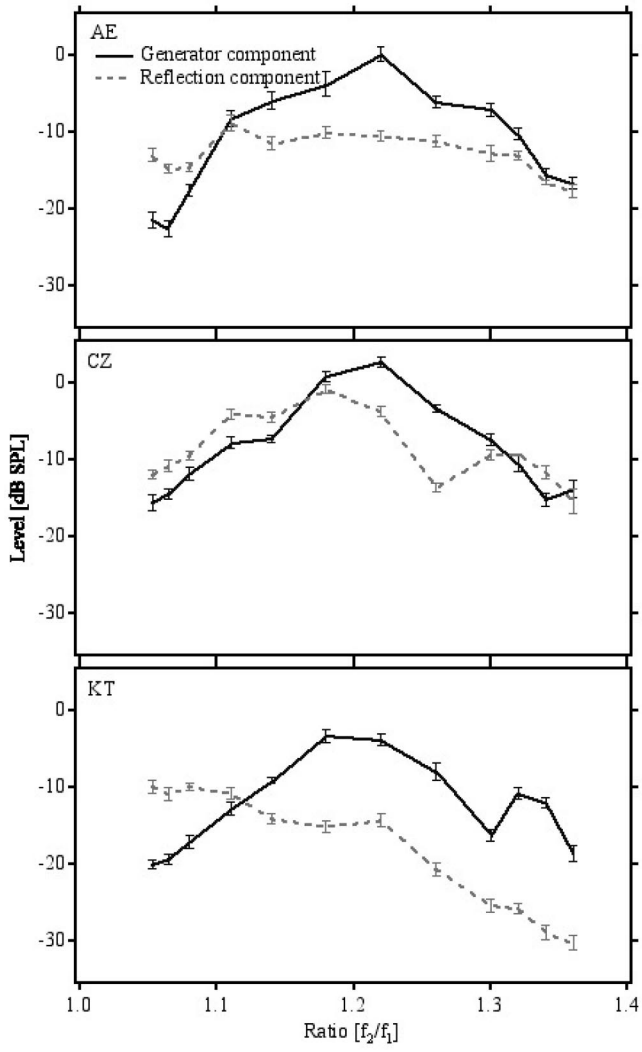


FIG. 3. Component magnitudes for subjects AE, CZ and KT with changing stimulus-frequency ratio. Data panels are marked with the subjects' initials. Solid and dashed lines represent A_{gen} and A_{ref}, respectively. Data used for these plots do not include those obtained with stimulus levels of 45 dB SPL. The error bars signify ± 1 standard error.

higher than A_{gen} at narrow ratios and the A_{gen} exhibits a bandpass characteristic in all subjects. A_{ref} is bandpass for subjects AE and CZ, but low pass for subject KT. Some additional differences between subjects can be observed by comparing the three panels of Fig. 3. In subject CZ, A_{ref} is more peaked compared to the other subjects. This results in a shift in the ratio at which the two traces cross over. While the crossover occurs around 1.11 for subjects AE (top panel) and KT (bottom panel), the crossover ratio is around 1.14 for subject CZ (middle panel). A_{gen} and A_{ref} show relative equivalence at high ratios for subjects AE and CZ, in keeping with the general trend seen in Fig. 2. While A_{ref} is larger than A_{gen} in subject CZ at these ratios ($\geqslant 1.3$), A_{gen} remains substantially larger than A_{ref} at high ratios for subject KT. A comparison of the three panels of Fig. 3 also reveals relative equivalence in the pattern and level of A_{gen} between the subjects. However, A_{ref} is very different for the three subjects, with A_{ref} being largest for subject CZ, followed by subject AE and KT.

A_{gen} and A_{ref} for different stimulus levels are presented in Fig. 4. Data for equal stimulus levels are presented in the left column, while the panels in the right column display data for unequal stimulus levels ($L_1 = 65$ and $L_2 = 45$, 50, 55, 60, and 65 dB SPL). While the general trends observed in the previous two figures are present, some significant variations with the stimulus level are noticable. The differences between A_{gen} and A_{ref} are less pronounced across the entire ratio range for stimulus levels of 65/55 and 65/60. Both A_{gen} and A_{ref} appear less peaked for stimulus levels of 45/45. Finally, A_{ref} remains consistently higher than A_{gen} across the entire ratio range for the lowest stimulus levels used (45 dB SPL). In contrast, little change is observed in A_{ref} at the highest stimulus levels used (75 dB SPL), with A_{gen} remaining higher than A_{ref} at wide ratios.

IV. DISCUSSION

By separating the ear-canal DPOAE signal into its two major components, we have been able not only to investigate the relative levels of the two components, but also to better

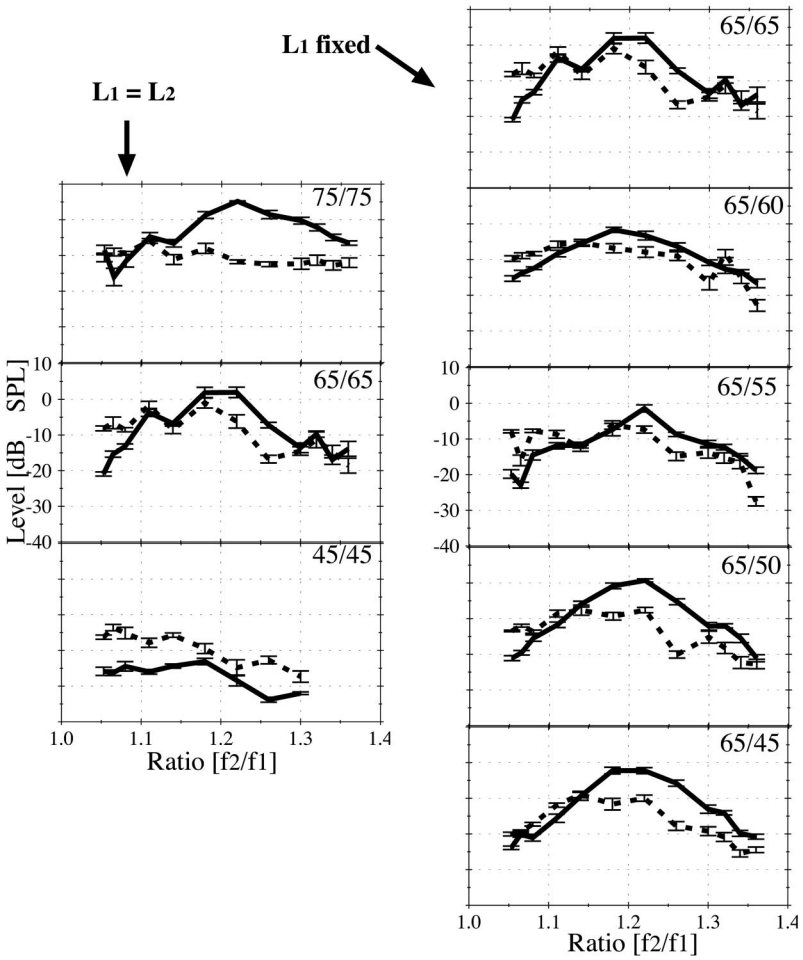


FIG. 4. A_{gen} and A_{ref} for different stimulus-level combinations as a function of the changing stimulus-frequency ratio. Solid and dashed lines represent A_{gen} and A_{ref}, respectively. The error bars signify ± 1 standard error.

understand the properties of each component. While the generator component (A_{gen}) when plotted as a function of the stimulus-frequency ratio showed a distinct bandpass shape with a peak near $f_1/f_2 \approx 1.2$ for all stimulus levels except 45/45 dB SPL, the pattern for the reflection component (A_{ref}) was more variable with the stimulus level and across ears. One of the most striking characteristics of the results is the stability of the generator component and the variability of the reflection component across ears. The influence of stimulus level on absolute and relative component levels is also found to be significant. We begin with a discussion of our data in the context of previous DPOAE publications and then explore the relationship of these data with basilar membrane mechanics. We conclude with some observations about the choice of stimulus parameters for the current clinical applications of DPOAEs.

A. Comparison with previous DPOAE results

The pattern we see in A_{gen} is similar to previous reports of the magnitude of the ear-canal DPOAE signal, when plotted as a function of the stimulus-frequency ratio (Harris *et al.*, 1989; Gaskill and Brown, 1990; O'Mahoney and Kemp, 1995; Stover *et al.*, 1999). Another notable feature in these reports was the presence of a notch in the DPOAE level around a stimulus-frequency ratio of 1.1 in human ears (e.g., Stover *et al.*, 1999). This notch has been interpreted as being due to the cancellation between the two DPOAE components when they are roughly equal in magnitude but almost 180° out of phase. In the few studies where amplitudes of the generator and reflection components, A_{gen} and A_{ref}, respectively, were extracted, the former exhibited bandpass behavior as a function of f_2/f_1 and the latter a low-pass or asymmetric bandpass behavior (Knight and Kemp, 1999, 2000, 2001). A_{ref} was $>A_{gen}$ for $f_2/f_1 < 1.1$ in humans (Knight and Kemp, 1999, 2000, 2001) and ≈ 1.2 in guinea pigs (Schneider *et al.*, 2003).

Before making more in-depth comparisons of our data with previous studies, some significant differences in methodology should be pointed out. While our recordings were made using repeated measures with f_2/f_1 fixed for each recording, Schneider *et al.* (2003) used a fixed-f_2 paradigm and estimated the relative dominance of the two components from group delay. Similarly, Konrad-Martin *et al.* (2001) also used a fixed-f_2 paradigm but extracted DPOAE component levels using an IFFT algorithm from the ratio function, resulting in data collapsed across frequency as well as stimulus-frequency ratio. This methodology served the purpose of their work—the examination of DPOAE component levels as a function of the stimulus level; but poses problems in direct comparison with our data. Perhaps most similar to our work in DPOAE-recording methodology are the reports by Knight and Kemp (2000, 2001), who also used IFFT analyses of results from fixed-ratio sweeps. However, signifi-

cant differences in stimulus levels used and the implementation of time-domain filters exist even here. Knight and Kemp use a narrow time window to isolate the (early latency) generator component and associate the reflection component with the accumulated signal that occurs after the early-latency one. In contrast, we use narrow time windows to isolate both the generator and the shortest-latency reflection components. This difference in approach influences the estimate of the reflection component significantly, since the wide time window used by Knight and Kemp allows the inclusion of secondary reflection components due to multiple internal reflections between the stapes and the DP CF region (Dhar *et al.*, 2002). The behavior of A_{gen}, observed here to be similar across subjects, is in close agreement with Knight and Kemp. However, we report much greater variability in the behavior of A_{ref} across subjects and demonstrate interesting effects of stimulus level on DPOAE components that have not been reported before.

Our results indicate a significant influence of stimulus level on the absolute and relative levels of the two components. Although Konrad-Martin *et al.* (2001) and Mauermann and Kollmeier (2004) have examined the effect of stimulus level on DPOAE component behavior in the past, it is difficult to isolate the effect of the stimulus-frequency ratio from that of the stimulus level in these studies. Both these groups use the ''scissor'' approach where the level difference between the stimulus tones is governed by a linear regression model (Kummer *et al.*, 1998; Janssen *et al.*, 1998), the difference decreasing with an increasing overall level. Additionally, Konrad-Martin *et al.* (2001) also report data recorded with a fixed level difference ($L_1-L_2=10$). Knight and Kemp (2000, 2001) used stimulus levels of 75 dB SPL only. In contrast, we have reported results for both equal- and unequal-level stimulus tones over a wide range (45–75 dB SPL). Our results are consistent with previous reports in that the reflection component is shown to be dominant at low stimulus levels only. Furthermore, we have also shown the dominance of A_{ref} to be consistent across the entire range of stimulus-frequency ratios tested at these levels.

Our results are loosely consistent with the previous measurements cited above for $f_2/f_1<1.3$ (Fig. 2). However, they differ from those of Knight and Kemp for $f_2/f_1>1.3$ in that A_{gen} and A_{ref} do not ''roll off'' together [see Fig. 5 of Knight and Kemp (2001)]. The relative behavior of A_{gen} and A_{ref} at these ratios is heavily dependent on the subject in our data (Fig. 3). A_{ref} is significantly lower than A_{gen} for subject KT, and the two component levels are more or less equivalent, with A_{gen} being slightly larger for subject AE. In subject CZ, A_{ref} is larger than A_{gen} at ratios >1.3. Thus, we see great variability in A_{ref} across subjects at these ratios. A_{gen} exhibits a bandpass behavior—with symmetric behavior on both sides of a peak around a ratio of 1.2, while the behavior of A_{ref} in most cases is asymmetric if bandpass or even low pass, with very little rolloff at narrow ratios. It is also noteworthy that our results are in close agreement with that obtained by Knight and Kemp, when data recorded with equivalent stimulus levels (e.g., 75/75 dB SPL) are compared.

The time-intensive experiments necessary to produce data on DPOAE components has limited the number of participant ears in our as well as previous reports. It is not surprising then that the variety of behavior in A_{ref} has gone unreported in the past. Our purely fortuitous sampling of ears appears to show greater diversity in this area as compared to previous subject pools. It is curious that the rank ordering observed in A_{ref} is consistent with other measures such as hearing thresholds and fine-structure depth. Although all three subjects demonstrated hearing sensitivity within normal limits, CZ had the best thresholds (coarsely measured using 5 dB steps and a clinical protocol), followed by AE and then KT. Fine-structure depth was the greatest in CZ and least in KT.

B. Comparison with psychophysical and physiological results

Insight into the behavior of DPOAE components may also be obtained by examining psychophysical and physiological data. The DP at $2f_1-f_2$, often referred to as the *cubic distortion tone* (CDT) in the physiological literature, has been examined by recording basilar membrane vibration at the CF place for $2f_1-f_2$ (see Robles and Ruggero, 2001, for a recent review). Termed the *cubic distortion tone* or the *combination tone* at $2f_1-f_2$ in the psychophysics literature, this DP has also been studied using cancellation as well as loudness-matching techniques (see Plomp, 1976). Recordings of basilar-membrane motion of the CDT have been reported from guinea pig [basal turn and hook, Cooper and Rhode (1997)] and chinchilla [apical turn, Cooper and Rhode (1997); basal turn Robles *et al.* (1997)] ears. The results depend on the stimulus level as well as the location of the preparation along the length of the cochlea. General observations from these data include a plateau in DP level as a function of increasing stimulus-frequency ratio up to a ''corner ratio'' and a sharp rolloff as stimulus-frequency ratio is increased beyond this critical ratio. The ''corner ratio'' is approximately dependent on stimulus level and the location in the cochlea. On occasion, a rolloff in the DP level as the stimulus-frequency ratio approaches 1 is observed. This rolloff is perhaps most-prominent in the hook and base regions as well as for higher stimulus levels (Cooper and Rhode, 1997). The reader is directed to Figures 4g–i of Cooper and Rhode (1997) and 6 and 7 of Robles *et al.* (1997). Note that these figures appear contrary at first glance, with the discrepancy resolved by accounting for the fact that Cooper and Rhode (1997) display the level of the DP relative to that of the lower-frequency stimulus tone. The cutoff ratio was highest and the slope the least steep for the recordings from the apical turn of the chinchilla cochlea.

The results from psychophysical experiments appear to be largely dependent on the method used for an estimation of the combination tone. A monotonic low-pass shape is observed with the use of the cancellation method (Goldstein, 1967), where the stimulus and cancellation tones are presented simultaneously. However, with the use of a nonsimultaneous pulsation technique, the combination tone appears to display an asymmetric bandpass shape (Smoorenburg, 1974; Siegel and Borneman, 1999), much akin to that observed in DPOAE and physiological data.

Prior to the emergence of DPOAE-component data, the discordance between the magnitudes of DPOAEs and basilar membrane motion as a function of the stimulus-frequency ratio was a source of concern. Indeed, Robles *et al.* (1997) contrast the lack of reduction in the level of $2f_1-f_2$ as f_2/f_1 approaches 1 in basilar-membrane measurements with that observed for DPOAEs. However, the more appropriate comparison of the DPOAE level with physiological or psychophysical data is made using A_{ref} rather than the ear-canal DPOAE level. Once the appropriate comparison is made, data from the basilar-membrane, psychophysical, and OAE domains appear to be in approximate qualitative agreement. It should be noted that to expect exact correspondence between data from these three domains is unrealistic due to differences in experimental conditions, and perhaps more importantly, the actual variables being measured. Whether the CDT measured directly on the basilar membrane, or the combination tone measured psychophysically, is equivalent to the reflection component of the DPOAE at $2f_1-f_2$ is an interesting question in itself. Even after the simplifying assumption of a transparent middle ear is made, the equivalence between these quantities would depend on the exact nature of the mechanism by which the incident DPOAE energy is ''turned around'' at the DPOAE-CF region. For example, Kemp's recent proposal of dichotomous *active* and *passive* reflectance mechanisms (Kemp, 2002) would alter whether the reflection term is ''accounted for'' in physiological and psychophysical measurements. However, it can be said with some confidence that measurements in all three domains represent some common aspects of peripheral auditory physiology.

C. Component magnitudes and basilar membrane mechanics

We discuss here the behaviors of both A_{gen} and A_{ref} in terms of cochlear models of DPOAEs [see Sec. V of Talmadge *et al.* (1998); and Shera (2003) for much greater model detail]. In the following greatly simplified discussion, we ignore details such as mutual suppression of the stimulus frequency waves when f_2/f_1 is close to 1, and focus on other factors that determine the gross variation of A_{gen} and A_{ref} with stimulus frequency ratio. A comparison with data would suggest that mutual suppression of the stimulus frequency waves mostly influence the behavior of A_{ref} for ratios close to 1.

For the case of a simple cubic nonlinearity, A_{gen} will be approximately proportional to the magnitude of an integral over the $2f_1-f_2$ sources. The integrand of the integral is proportional to the square of the apical-moving f_1 cochlear wave times the complex conjugate of the apical-moving f_2 cochlear wave times the $2f_1-f_2$ apical-moving cochlear wave [see Eqs. (130) through (136) in Talmadge *et al.* (1998)]. The expression for A_{ref} involves the magnitude of the product of R_a (the apical reflectance that results from coherent reflection of a cochlear wave due to cochlear roughness around the DP tonotopic site) times an integration over sources that is essentially the same as the one for A_{gen}, except that the basal-moving DP wave replaces the apical-moving one in the integrand. For large f_2/f_1, both A_{gen} and A_{ref} decrease with increasing frequency ratio largely due to the decrease of overlap between the activity patterns of the two stimulus waves.

When f_2/f_1 is close to 1, the magnitude of the source integrations for A_{gen} and A_{ref} are mainly determined by contributions from around the f_2 tonotopic site ($\hat{x}_2$). The frequency ratio that will give the largest value of A_{gen} will then be approximately the value that minimizes the variation of the phase of the integrand with distance x along the BM in the region of $x \approx \hat{x}_2$ (the principle of the stationary phase). This phase condition can be simply expressed as (Talmadge *et al.*, 1998):

$$\hat{k}-2k(\hat{x}_2,f_1)-k(\hat{x}_2,2f_1-f_2)=0, \tag{2}$$

where $\hat{k}$ is the wave number of a cochlear wave in its peak region (constant in a scale-invariant cochlea) and $k(x,f)$ is the wave number of a cochlear wave of frequency f at location x. The dependences of $k(\hat{x}_2,f_1)$ and $k(\hat{x}_2,2f_1-f_2)$ on f_2/f_1 are estimated in Talmadge *et al.* (1998), where it is also shown that

$$A_{\text{gen}} \approx \cdots \exp\{-(\sigma_x^2/2)[\hat{k}-2k(\hat{x}_2,f_1)-k(\hat{x}_2,2f_1-f_2)]^2\}, \tag{3}$$

where σ_x is a measure of the spatial width of the peak of the cochlear activity pattern. As f_2/f_1 increases from 1, the expression on the left side of Eq. (2) starts out negative and increases through zero. These considerations then imply that A_{gen} as a function of f_2/f_1 has bandpass behavior, in agreement with data presented previously as well as here.

Now consider A_{ref}. Using an approach similar to the one just discussed, it can be shown that

$$A_{\text{ref}} \approx \cdots |R_a| \exp\{-(\sigma_x^2/2)[\hat{k}-2k(\hat{x}_2,f_1) + k(\hat{x}_2,2f_1-f_2)]^2\}. \tag{4}$$

As f_2/f_1 increases from 1, the quantity,

$$\hat{k}-2k(\hat{x}_2,f_1)+k(\hat{x}_2,2f_1-f_2), \tag{5}$$

in Eq. (4) can be shown to start out positive or negative and increase with increasing ratio. If the value is initially positive, A_{ref} will monotonically decrease with increasing f_2/f_1, resulting in a low-pass shape (similar to that observed for subject KT in Fig. 3). On the other hand, when the value is initially negative, it is still greater than the quantity on the left-hand side of Eq. (2). In this case, A_{ref} will assume an asymmetric bandpass shape with the peak at a ratio less than the peak ratio for A_{gen} (similar to that observed for subject CZ in Fig. 3). Additionally, this relationship leads to the prediction of $A_{\text{ref}} > A_{\text{gen}}$ as f_2/f_1 increases from 1, given a reasonable magnitude of R_a. A_{ref} is predicted to be equal to A_{gen} (and possibly give rise to a notch behavior in the DPOAE amplitude under the requisite phase-cancellation conditions) for a ratio (typically around 1.1 in our data) smaller than the peak ratio for A_{gen} (typically around 1.2), in agreement with the data of Figs. 2, 3, 4.

From Eqs. (3) and (4), we see that

$$A_{\text{ref}}/A_{\text{gen}} \approx |Ra| \exp\{(2\sigma_x^2)k(\hat{x}_2,2f_1-f_2) \times [\hat{k}-2k(\hat{x}_2,f_1)]\}. \tag{6}$$

It should be emphasized that R_a, according to the coherent reflection model (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera, 2003), depends mainly on the sharpness of tuning of the BM around the DP CF region and the strength of cochlear inhomogeneities around that region. R_a will generally be significant only if both of these factors are present. Thus R_a is expected to fluctuate significantly across healthy ears due to differences in the degree of cochlear roughness across such ears. It is therefore not surprising that the relative levels of A_{gen} and A_{ref} are found to change from ear to ear for $f_2/f_1 > 1.3$ and across frequency within a given ear (see Fig. 1).

In sum $A_{\text{ref}} > A_{\text{gen}}$ at narrow ratios, overcoming the suppressive effects of the stimulus tones on R_a. This observation is in contrast to the assumption of bidirectional distribution of energy from the generation region (e.g., Allen and Fahey, 1992). Thus, Shera's (2003) assertion that phase appears to ''win'' in the battle between amplitude and phase in the overlap region can perhaps be confidently extended; not only does phase win, it appears to win by a landslide. The behavior of A_{ref} at wide stimulus-frequency ratios is strongly influenced by R_a. We provide two lines of evidence in support of R_a being the driving force behind the behavior of A_{ref} at wide ratios. First, with an (assumptive) correlation between cochlear health and reflectance, we argue that the rank ordering of A_{ref}, hearing thresholds, and fine-structure depth in our three subjects is not coincidental but due to varying degrees of reflectance in these ears. Second, the influence of R_a on A_{ref} is evident in Fig. 4 where A_{ref} remains larger than A_{gen} for all stimulus-frequency ratios for stimulus levels of 45 dB SPL, due to reduced suppression by the lower-level stimulus tones. The influence of stimulus level on A_{ref} has also been demonstrated elsewhere (Mauermann and Kollmeier, 2004; Konrad-Martin *et al.*, 2001).

We have thus seen that most of the characteristic features of A_{gen} and A_{ref} that we find are at least qualitatively described by cochlear models that incorporate a nonlinear generation process around the f_2 CF region combined with the coherent reflection of the initial apical-moving DP component.

D. Clinical implications

We conclude by making some observations about the relationship between A_{gen} and A_{ref} when stimulus levels (L_1/L_2 in dB SPL) of 65/50 and 65/55 are used. A stimulus-level combination of 65/55 at a stimulus-frequency ratio of 1.2 is often used in clinical applications of DPOAEs. There has been some recent work speculating that DPOAE fine structure is (at least partially) responsible for the failure of DPOAEs in predicting hearing thresholds with accuracy (Shaffer *et al.*, 2003; Shera, 2004; Dhar and Shaffer, 2004). If this speculation was valid and the goal of a clinical protocol was to record DPOAEs with the minimum possible fine structure, stimulus conditions that lead to the biggest difference between A_{ref} and A_{gen} should be chosen. More specifically, the ideal stimulus conditions should bias the ear-canal DPOAE signal toward A_{gen} to the greatest extent possible, if the goal is to monitor cochlear health around the f_2 CF region. Our data appear to suggest that stimulus-frequency ratios around 1.22 satisfy these demands to a great extent, with the exception that a stimulus-frequency ratio of 1.26 leads to similar if not greater bias toward the generator component when stimulus levels of 65/50 are chosen. Considerations such as a good signal-to-noise ratio and the separation between normal and impaired populations also influence the choice of stimulus conditions in a clinical protocol. However, our results indicate that the current choice of stimulus parameters in clinical protocols mostly satisfies the consideration of relative component levels.

ACKNOWLEDGMENTS

Part of this work was completed when all four authors were affiliated with Purdue University. This research was supported in part by the NIH/NIDCD Grant No. R29 DC03094 to CLT and R03 DC0005692 and a grant from the DRF to SD. We thank Professors Christopher Shera, Paul Fahey, Jonathan Siegel, and Robert Withnell for their thoughtful comments on an earlier version of the paper. Thanks are also due to Professor Brenda Lonsbury-Martin and two anonymous reviewers. Feedback received from them made this a better manuscript.

Allen, J. B., and Fahey, P. F. (**1992**). ''Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane,'' J. Acoust. Soc. Am. **92**, 178–188.

Cooper, N. P., and Rhode, W. S. (**1997**). ''Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea,'' J. Neurophysiol. **78**, 261–270.

Dhar, S., and Shaffer, L. A. (**2004**). ''Effects of a suppressor tone on distortion product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-grams,'' Ear Hear. **25**, 573–585.

Dhar, S., Talmadge, C., Long, G., and Tubis, A. (**2002**). ''Multiple internal reflections in the cochlea and their effect on DPOAE fine structure,'' J. Acoust. Soc. Am. **112**, 2882–2897.

Gaskill, S. A., and Brown, A. M. (**1990**). ''The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity,'' J. Acoust. Soc. Am. **88**, 821–839.

Goldstein, J. L. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.

Greenwood, D. D. (**1990**). ''A cochlear frequency position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). ''Acoustic distortion products in humans: Systematic changes in amplitude as a function of f_2/f_1 ratio,'' J. Acoust. Soc. Am. **85**, 220–229.

Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). ''Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1-f_2$ removes DP-gram fine structure—evidence for a secondary generator,'' J. Acoust. Soc. Am. **103**, 1527–1531.

Janssen, T., Kummer, P., and Arnold, W. (**1998**). ''Growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission in tinnitus,'' J. Acoust. Soc. Am. **103**, 3418–3429.

Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: A test of the two-mechanism model for dpoae generation,'' J. Acoust. Soc. Am. **109**, 622–637.

Kemp, D. T. (**1986**). ''Otoacoustic emissions, travelling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.

Kemp, D. T. (**2002**). *Exploring Cochlear Status With Otoacoustic Emissions: The Potential For New Clinical Application*, edited by M. Robinette and T. Glattke, Thieme, New York.

Kim, D. O. (**1980**). ''Cochlear mechanics: Implications of electrophysiological and acoustical observations,'' Hear. Res. **2**, 297–317.

Knight, R. D., and Kemp, D. T. (**1999**). ''Relationship between DPOAE and TEOAE characteristics,'' J. Acoust. Soc. Am. **106**, 1420–1435.

Knight, R. D., and Kemp, D. T. (**2000**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1, f_2 area study,'' J. Acoust. Soc. Am. **107**, 457–473.

Knight, R. D., and Kemp, D. T. (**2001**). ''Wave and place fixed DPOAE maps of the human ear,'' J. Acoust. Soc. Am. **109**, 1513–1525.

Konrad-Martin, D., Neely, S., Keefe, D., Dorn, P., Cyr, E., and Gorga, M. (**2002**). ''Sources of DPOAES revealed by suppression experiments, inverse fast Fourier transforms, and SFOAES in impaired ears,'' J. Acoust. Soc. Am. **111**, 1800–1809.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). ''Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears,'' J. Acoust. Soc. Am. **109**, 2862–2879.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f_1-f_2$ distortion product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.

Long, G. R., Shaffer, L. A., Dhar, S., and Talmadge, C. L. (**1999**). ''Cross species comparison of otoacoustic fine-structure,'' in *Recent Developments in Auditory Mechanics* (World Scientific, Singapore), pp. 367–373.

Mauermann, M., and Kollmeier, B. (**2004**). ''Distortion product otoacoustic emission (DPOAE) input/ouput functions and the influence of the second dpoae source,'' J. Acoust. Soc. Am. **116**, 2199–2212.

Mauermann, M., Uppenkamp, S., Hengel, P. W. J. van, and Kollmeier, B. (**1999**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss,'' J. Acoust. Soc. Am. **106**, 3484–3491.

O'Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emission delay measurement in humans,'' J. Acoust. Soc. Am. **97**, 3721–3735.

Plomp, R. (**1976**). *Aspects of Tone Sensation*, Academic, London.

Ren, T. (**2004**). ''Reverse propagation of sound in the gerbil cochlea,'' Nat. Neurosci. **7**, 333–334.

Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). ''Two-tone distortion on the basilar membrane of the chinchilla cochlea,'' J. Neurophysiol. **77**, 2385–2399.

Ruggero, M. A. (**2004**). ''Comparison of group delays of $2f_1-f_2$ distortion product otoacoustic emissions and cochlear travel times,'' ARLO **5**, 143–147.

Schneider, S., Prijs, V. F., and Schoonhoven, R. (**2003**). ''Amplitude and phase of distortion product otoacoustic emissions in the guinea pig in an (f_1,f_2) area study,'' J. Acoust. Soc. Am. **113**, 3285–3296.

Shaffer, L., Withnell, R., Dhar, S., Lilly, D., Goodman, S., and Harmon, K. (**2003**). ''Sources and mechanisms of dpoae generation: Implications for the prediction of auditory sensitivity,'' Ear Hear. **25**, 367–379.

Shera, C. (**2004**). ''Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions,'' Ear Hear. **25**, 86–97.

Shera, C. A. (**2003**). ''Wave interference in the generation of reflection and distortion-source emissions,'' in *Biophysics of the Cochlea: From Molecules to Models* (World Scientific, Singapore), pp. 311–320.

Siegel, J. H., and Borneman, A. L. (**1999**). ''Comparable measures of perceptual combination tones and distortion-product otoacoustic emissions,'' in ''Abstracts of the association for research in otolaryngology.''

Smoorenburg, G. (**1974**). ''On the mechanism of combination tone generation and lateral inhibition in hearing,'' in *Psychological Models and Physiological Facts in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, Berlin), pp. 1–10.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). ''Latency and multiple sources of distortion product emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1999**). ''Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears,'' J. Acoust. Soc. Am. **106**, 2669–2678.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing theshold fine structures in humans,'' J. Acoust. Soc. Am. **104**, 1517–1543.

Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). ''Generation of DPOAES in the guinea pig,'' Hear. Res. **178**, 106–117.

Zweig, G., and Shera, C. A. (**1995**). ''The origins of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

A comparison of various nonlinear models of cochlear compression

James M. Harte[a] and Stephen J. Elliott
Institute of Sound and Vibration Research, University of Southampton, University Road, Southampton, Hampshire, SO17 1BJ, United Kingdom

Henry J. Rice
Department of Mechanical and Manufacturing Engineering, Trinity College, Dublin 2, Ireland

(Received 20 December 2004; revised 18 March 2005; accepted 22 March 2005)

The vibration response of the basilar membrane in the cochlea to sinusoidal excitation displays a compressive nonlinearity, conventionally described using an input-output level curve. This displays a slope of 1 dB/dB at low levels and a slope $m<1$ dB/dB at higher levels. Two classes of nonlinear systems have been considered as models of this response, one class with static power-law nonlinearity and one class with level-dependent properties (using either an automatic gain control or a Van der Pol oscillator). By carefully choosing their parameters, it is shown that all models can produce level curves that are similar to those measured on the basilar membrane. The models differ, however, in their distortion properties, transient responses, and instantaneous input-output characteristics. The static nonlinearities have a single-valued instantaneous characteristic that is the same at all input levels. The level-dependent systems are multi-valued with an almost linear characteristic, for a given amplitude of excitation, whose slope varies with the excitation level. This observation suggests that historical attempts to use functional modeling (i.e., Wiener of Volterra series) may be ill founded, as these methods are unable to represent level-dependent nonlinear systems with multi-valued characteristics of this kind. [DOI: 10.1121/1.1906059]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.25.Ts [BLM] Pages: 3777–3786

I. INTRODUCTION

The human auditory system is known to compress a dynamic range of about 120 dB, when considering rms fluctuations in air pressure, to around 60 dB for neural coding (Allen and Neely, 1992). This suggests the presence of a nonlinear mechanism in the ear. One source of these nonlinearities is commonly believed to be a feedback loop within the cochlea, often described as the ''active process.'' This was first postulated by Gold (1948), whose theory stated that to account for the high frequency selectivity the filters in the cochlear must be active. This mechanism sharpens the response of the auditory filters, with the potential to oscillate if too much energy is fed into them. As well as compressing the dynamic range of the cochlear response, this active process is also believed to give rise to otoacoustic emissions (OAEs), as first discovered by Kemp (1978).

When the active cochlea is driven sinusoidally, the level of the basilar membrane (BM) vibration grows in a nonlinear or compressive way with sound pressure level. These results have been reported in the literature, for example by Rhode (1971), Sellick *et al.* (1982), Robles *et al.* (1986), Nuttall and Dolan (1996), and Rhode and Recio (2000). A commonly accepted observation is that at low levels a 1-dB increase in sound pressure level produces a 1-dB increase in BM vibration level (indicating a linear response), but for sound pressure levels above about 30 dB (relative to 20 μPa), a 1-dB increase in sound pressure level only produces a much smaller increase in BM vibration level, typically less than 0.5 dB/dB. In fact Ruggero *et al.* (1997) have observed level curves of as little as 0.2 dB/dB.

The physiological data available in the literature were obtained using invasive methods predominantly from animals such as chinchillas, guinea pigs, and squirrel monkeys. This data is not available in humans, other than from cadavers known not to retain the compressive response of the active process. Hence various psychoacoustic measures have been pursued to investigate the compression in the cochlea [see for example Moore *et al.* (1999)]. However, these approaches are hard to use for gainful system identification purposes.

OAEs have also been shown to exhibit similar compressive characteristics to those present in the mechanical response of the cochlea, and as such have the potential to be used as a noninvasive method for cochlear research. Kemp (1978) first reported the compression in the level curve, and linked it to the essential amplitude nonlinearity of cochlear mechanics. The level curves for tone-burst, distortion product, and transient-evoked OAEs have been reported by Norton and Neely (1987), Dorn *et al.* (2001), Zwicker (1983), Zwicker and Schloth (1984), and Prieve *et al.* (1996). They were all shown to exhibit a linear region for low excitation levels and a compressive region for higher levels. The slopes of these level curves varied between 0.2 and 0.6 dB/dB.

The first purpose of this investigation is to examine the properties of various historically used models of nonlinear

[a]Author to whom correspondence should be addressed. Electronic mail: jmh@isvr.soton.ac.uk

cochlear function, and demonstrate how they can all produce input-output level curves with the same characteristics as those measured in the BM and OAEs. We will also investigate various other properties of these models, in particular the amount of distortion for tonal excitation, the transient responses, and the instantaneous input-output characteristics, and show how these differ for the various models. These simple nonlinear models are entirely local, however, and although they can reproduce the compression curves for steady-state sinusoidal signals, a distributed model of the mechanics along the cochlea would be required to allow meaningful comparison with more detailed physiological measurements.

There exists a large body of literature on the various attempts to model the nonlinearities in the cochlea. It is not the purpose of this paper to exhaustively review this literature but instead to examine the simple generic types of nonlinearity that are widely used. For a more general review of cochlear modeling the reader is referred to de Boer (1996). Static nonlinear models are considered in Sec. II. In Sec. III two level-dependent systems are considered, the first a simple automatic gain control (AGC), the second, a system based on Van der Pol's equation. Van der Pol's equation can display a wide variety of dynamic behaviors including limit-cycle oscillations, whose properties have striking similarities to those of spontaneous otoacoustic emissions. The distortion properties of all the nonlinear systems will be discussed alongside the input-output level curves. Section IV compares the responses to the different models due to sinusoidal excitation, discussing their transient responses, and instantaneous input-output characteristics.

II. STATIC NONLINEAR MODELS

Static or memoryless nonlinearities are defined such that the current output signal, $y(t)$, is a function only of the current input signal, $x(t)$, i.e.,

$$y(t)=f(x(t)). \tag{1}$$

Static nonlinearities have been used extensively in the analysis of the human auditory system. See, for example, Pfeiffer (1970) in his attempts to model the auditory pathways leading to the generation of the cochlear microphonic (CM), van Dijk *et al.* (1994, 1997a,b) used a similar approach to Pfeiffer except with the intention of studying the output from a single auditory nerve pierced with an electrode, rather than the CM, Lopez-Poveda and Meddis (2001) and Meddis *et al.* (2001) have used a static nonlinear element in a phenomenological model of cochlear compression. Bian *et al.* (2002) have also used instantaneous nonlinearities to explain the effects of low frequency modulation on distortion product otoacoustic emissions, although in a later paper (Bian *et al.*, 2004) a hysteretic element was also included.

A simple example of a static nonlinear system is a power-law nonlinearity; for example,

$$y(t)=|x(t)|^p \operatorname{sgn}[x(t)], \tag{2}$$

where the compressive effect is ensured if the power, p, is less than unity. The signum function, defined by

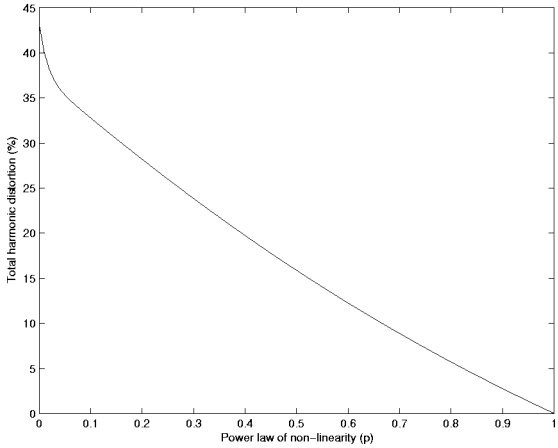


FIG. 1. The variation with exponent p of the percentage total harmonic distortion for sinusoidal excitation of the power-law nonlinearity.

$$\operatorname{sgn}[x(t)]=\begin{cases}-1 & \text{for } x<0,\\ +1 & \text{for } x\geqslant 0,\end{cases} \tag{3}$$

ensures asymmetry in the nonlinear characteristic. If such a system is excited by sinusoidal excitation of the form

$$x(t)=\sqrt{2}X\cos(\omega_0 t), \tag{4}$$

where X is the rms input amplitude and ω_0 is the angular excitation frequency. The output from such a system will be a periodic signal with a Fourier series in which the amplitude of the nth harmonic of the output is equal to (Elliott and Harte, 2003)

$$Y_n=a_n 2^{(p-1)/2}X^p, \tag{5}$$

where a_n is a level-independent constant. In particular, the fundamental component of the output is given by

$$Y_1=a_1 2^{(p-1)/2}X^p. \tag{6}$$

Obtaining the level curve for this power-law nonlinearity by plotting $20\log_{10}Y_1$, the log output fundamental, against the log excitation amplitude $20\log_{10}X$, will thus yield a slope of p dB/dB.

Varying the exponent between $p=0$, where the output signal is a square wave, and 1, where the output is a sinusoid, shows the transition from perfect compression to linearity. Total harmonic distortion is a simple measure of distortion, defined by the ratio of the power in the harmonics of the signal minus the fundamental to the total power in the signal:

$$\Xi(\%)=\left(\frac{\Sigma_{n=2}^{\infty}|Y_n|^2}{\Sigma_{n=1}^{\infty}|Y_n|^2}\right)^{1/2}\times 100\%, \tag{7}$$

where Y_n is the nth-harmonic component of the output time series. Figure 1 shows the variation with p of the total harmonic distortion. The total harmonic distortion of the static nonlinearity in Eq. (2) is about 44% when $p=0$, corresponding to the distortion of a square wave, falling to about 17% when $p=0.5$ and to zero when $p=1$, when the output is undistorted.

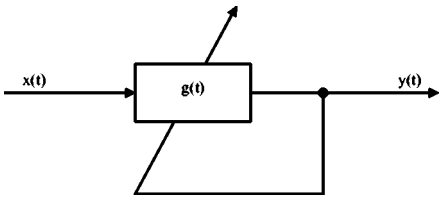


FIG. 2. Schematic for instantaneously acting automatic gain control system.

A. Input-scheduled power law

In order to incorporate the feature of a level curve with a dual slope (unity at low levels and less than unity for higher levels of excitation), it is necessary for the power law to change with the instantaneous value of the input signal, so that it may be written as $p(x)$, and the output signal is now given by

$$y(t)=|\alpha x(t)|^{p(x)}\,\mathrm{sgn}[x(t)], \tag{8}$$

where α is a constant. We will assume that the variation of the power with the instantaneous value of the input signal takes the form

$$p(x)=\tfrac{3}{4}-\tfrac{1}{4}\tanh(|x|-\varepsilon), \tag{9}$$

where ε is a threshold value for x, well below which $p(x)\approx 1$ and well above which $p(x)\approx 0.5$. This value for the exponent p was chosen to ensure a 0.5 dB/dB level curve slope at high input levels. A level curve slope of 0.5 dB/dB was chosen as an example slope of BM velocity observed in the literature. Other slopes have been noted as low as 0.2 dB/dB, however this is only an illustrative example and shall be used throughout this paper in all modeling efforts. The value of ε has been taken to be $10^{1.5}$ to ensure that the level curve changed slope at the appropriate input level of 30 dB.

The harmonic distortion of this nonlinearity is zero at low amplitudes and rises to about 17% at high amplitudes, as predicted for a power-law nonlinearity with an exponent varying between $0\leqslant p\leqslant 0.5$.

B. Output-scheduled power law

Another static nonlinear system considered is a power law scheduled on the instantaneous output signal. This can be considered as a limiting case of the automatic gain control model considered below, which acts instantaneously. Figure 2 shows the schematic of an instantaneous gain function, $g(t)$, scheduled by the instantaneous output signal $y(t)$. The system equation is given by

$$y(t)=g(t)x(t), \tag{10}$$

where $g(t)$ is the instantaneous gain. We will assume that $g(t)$ takes the form

$$g(t)=\frac{1}{|y(t)|+\varepsilon},\quad \varepsilon>0, \tag{11}$$

where ε is a constant. If $|y(t)|\gg\varepsilon$, the system will behave to give a compressive nonlinearity with a level curve of slope 0.5 dB/dB. If $|y(t)|\ll\varepsilon$, however, then the gain of the system is independent of the input and the slope of the input-output level curve will be 1 dB/dB. From this equation it is clear

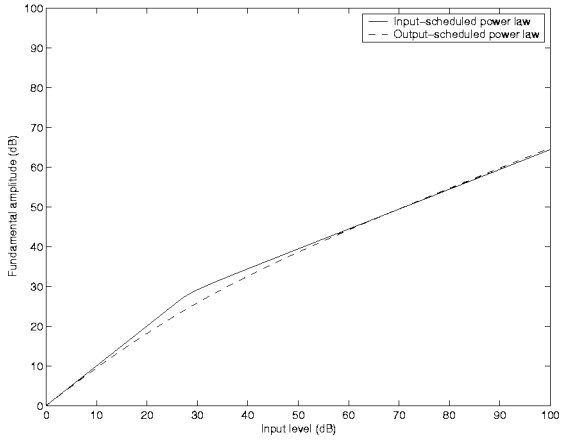


FIG. 3. The level curves for sinusoidal excitation of the two instantaneous nonlinear systems: the input-scheduled power law (solid line) and the output-scheduled power law (dashed line).

that the instantaneous gain $g(t)$ is always nonzero and positive. Inserting this into Eq. (10) and noting that $|y(t)|=y(t)\mathrm{sgn}[y(t)]$, a quadratic equation in $y(t)$ can be obtained, whose solution is given by (Harte, 2004)

$$y(t)=\frac{-\varepsilon\,\mathrm{sgn}[x(t)]\pm\sqrt{\varepsilon^2+4|x(t)|}}{2}. \tag{12}$$

By examining the limiting cases as $x(t)\to 0$ from $\pm\infty$, and for the specific case of $x(t)=0$, it is possible to show that the physically realistic solution for Eq. (12) is given by

$$y(t)=\tfrac{1}{2}(\sqrt{\varepsilon^2+4|x(t)|}-\varepsilon)\mathrm{sgn}[x(t)]. \tag{13}$$

Equation (13) can be considered an instantaneously acting automatic gain control, i.e., a static nonlinear system which is scheduled on the output $y(t)$. The harmonic distortion for this nonlinearity is very low for low amplitude inputs and rises to about 17% at high amplitudes, in a similar manner to the input-scheduled power-law nonlinearity.

C. Input-output level curves

The level of the fundamental component of the output response plotted against the level of the input signal gives the level curve of the system. Figure 3 shows the level curves computed from these two static nonlinearities. The solid curve and the dashed curves give the level curves for the input- and output-scheduled power laws, respectively, and have a slope of 1 dB/dB at low levels and of 0.5 dB/dB at higher levels. The level curves are asymptotically very similar and it may thus be difficult to choose between these competing models on the basis of these steady-state level curves.

III. DYNAMIC NONLINEAR MODELS

Two level-dependent nonlinear systems will be developed in this section, an automatic gain control and a model based on a variant of the Van der Pol oscillator.

A. Automatic gain control

Lyon (1990) introduced a model of the cochlea with explicit level-dependent gain (i.e., an automatic gain control

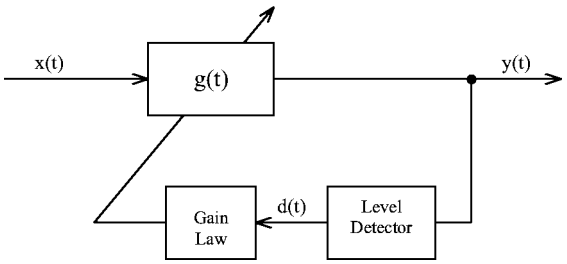


FIG. 4. Block diagram of a single automatic gain control whose gain depends on the detected output signal.

or AGC). Lyon and Mead (1988) developed an electronic cochlea using an analog implementation of AGC circuits. Kanis and de Boer (1993) in their ''quasilinear'' model use a variable impedance that depends on the level of the excitation pressure, and so this too constitutes a level-dependent model. Prior to these works Zwicker (1979, 1986) developed an analog electronic dynamic model of the cochlea. A feedback loop representing the contribution to local BM velocity of the outer hair cells was used, consisting of a point model with an amplifier and a saturating instantaneous nonlinear device, to produce the compression in the system. This form of system was also investigated by Yates (1990), who demonstrated that it can essentially be considered as a form of feedback AGC.

A schematic diagram of the feedback AGC (i.e., a system in which the gain depends on the level of the output) is shown in Fig. 4. The level of the output is measured using an output detector and used to control the instantaneous gain, $g(t)$, via a gain law. The AGC system will display a level curve that has a slope of 1 dB/dB up to a specified input level and a slope of m dB/dB at higher levels, if the gain law is of the form

$$g=\frac{1}{d^p+\varepsilon}, \tag{14}$$

where $p=(1-m)/m$ is the power or the exponent of the detected signal, ε is a constant, and d is the detector circuit output that ideally should be equal to the rms output amplitude Y. If $d\ll\varepsilon$, then the gain of the system is independent of input level and the slope of the input-output level curve will be 1 dB/dB. If $d\gg\varepsilon$, then the system will behave to give a compressive nonlinearity with a level curve of slope m dB/dB. Letting $m=0.5$, the transition from linearity at low levels to a compressive 0.5 dB/dB growth occurs after some threshold determined by the value of the parameter ε.

The detector used in the AGC is assumed to consist of an instantaneous squaring device, a low-pass filter, and an instantaneous square-root device. The low-pass filter is described in differential equation form by

$$u(t)=v(t)+T_{\mathrm{LPF}}\frac{dv(t)}{dt}, \tag{15}$$

where $u(t)$ and $v(t)$ are the input and output signals of the low-pass filter, and T_{LPF} is the time constant. The frequency response of this system is given by

$$H(j\omega)=\frac{\omega_{\mathrm{LPF}}}{\omega_{\mathrm{LPF}}+j\omega}, \tag{16}$$

where $\omega_{\mathrm{LPF}}=1/T_{\mathrm{LPF}}$. The overall transient response of the AGC is due to the low-pass filter's dynamic behavior in the level detector. It was shown (Elliott and Harte, 2003) that for small but abrupt changes in input level, the variation of the AGC gain with time is exponential, with time constant, T_{AGC}, which is half that of the low-pass filter T_{LPF}, so that

$$T_{\mathrm{AGC}}=\frac{1}{2\omega_{\mathrm{LPF}}}. \tag{17}$$

The sinusoidal input signal is assumed to be at the characteristic frequency for the BM site and to have an angular frequency of ω_c. Then its period is equal to

$$T_c=\frac{2\pi}{\omega_c}. \tag{18}$$

The normalized AGC time constant, given by the ratio of time constant of the AGC to the period of the excitation frequency, is thus defined to be

$$\chi=\frac{T_{\mathrm{AGC}}}{T_c}=\frac{\omega_c}{4\pi\omega_{\mathrm{LPF}}}. \tag{19}$$

If all the components of the low-pass filter's input are well above ω_{LPF}, then the low-pass filter will smooth out this signal, to give a constant, dc, value, which is then used to schedule the gain. In particular, if the input is a sinusoid with an angular frequency, ω_c, which is well above ω_{LPF}, then the detector signal, $d(t)$, will be equal to the rms output amplitude of the output signal:

$$d(t)=Y. \tag{20}$$

However, if the input frequency is not much larger than the cutoff frequency of the low-pass filter, then the output detector, $d(t)$, will have an alternating as well as a steady component. This will cause the gain of the AGC to vary during the course of a cycle of the input signal, thus distorting it and generating harmonics in the output signal.

The total harmonic distortion of such an AGC at high input levels was shown by Elliott and Harte (2003) to be a function of the ratio of the input frequency to the cutoff frequency of the detector, and hence the normalized time constant of the low-pass filter. If the normalized time constant χ is not too small, then the total harmonic distortion given as a percentage can be shown to be approximately equal to

$$\Xi(\%)\approx\frac{1}{\chi}. \tag{21}$$

Figure 5 shows the total harmonic distortion measured in simulations of the AGC for a sinusoidal input signal as a function of the normalized time constant χ. It can be seen that the theoretical predictions for the total harmonic distortion match the simulation results well for a normalized time constant above around 0.1, but cannot predict the more complicated behavior seen in the simulations for lower values of χ because the assumptions made on the derivation are no longer valid in this case.

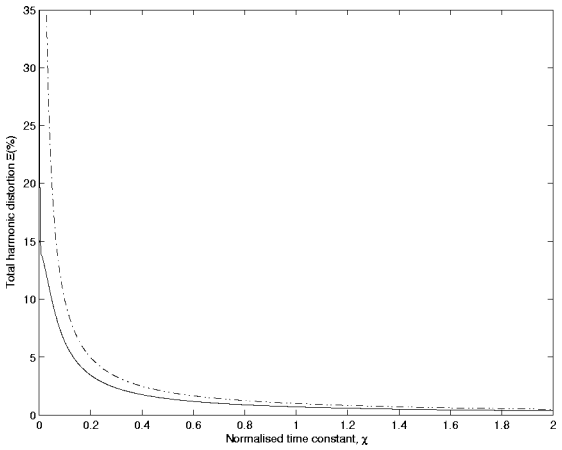


FIG. 5. Total harmonic distortion of the AGC for a sinusoidal input signal as a function of the normalized time constant, χ. The solid curve is the results of computer simulations and the dashed curve is the theoretical prediction, that $\Xi = 1/\chi$.

The total harmonic distortion of an AGC with a gain law of the form of Eq. (14) will vary with level in a similar manner to that of the static nonlinearity in Sec. II B, beginning essentially zero at low levels and rising to a constant value at input levels above about 30 dB, as will be illustrated below. For a given excitation frequency, the AGC time constant results in a compromise between low distortion (large χ) and fast response time (small χ).

B. Limit-cycle oscillator

For a dynamic, mass-spring-damper model of BM response, in which the damping is a function of the amplitude of response, the gain is implicitly level dependent. Such a model can be based on Van der Pol's equation, which can display a wide variety of dynamic behaviors including limit cycle oscillations, whose properties have striking similarities to those of spontaneous otoacoustic emissions, and is described by a ''Hopf-bifurcation.'' They operate from a stable equilibrium to a limit cycle oscillation of increasing amplitude, as the nonlinear damping parameter is varied. The introduction of such models is opening up a new approach to cochlear modeling, with the use of Hopf bifurcations to account for well-documented essential nonlinearities of the ear [see Eguíluz *et al.* (2000) and Camalet *et al.* (2000)], for example the compression of dynamic range (input-output level curve), sharper cochlear tuning for softer sounds, and generation of combination tones. There also exists a body of literature which argues for a limit cycle oscillator, such as a Van der Pol system, as a sufficient model for the qualitative features in auditory nonlinearity. See, for example, Choe *et al.* (2000), Jülicher and Prost (1997), Martin and Hudspeth (2001), and Ospeck *et al.* (2001).

Spontaneous otoacoustic emissions (SOAEs) are a particularly important subclass of OAEs. Rabinowitz and Widin (1984) presented an early report of some basic properties of SOAEs and their interactions with external tones and broadband noise stimuli. Bialek and Wit (1984), van Dijk and Wit (1990), and Talmadge *et al.* (1991) measured the statistical properties of SOAEs and compared them with the theory of nonlinear oscillators. Long *et al.* (1991) argue that many aspects of the interaction of SOAEs with external tones can be simulated by the behavior of a single driven Van der Pol oscillator. Murphy *et al.* (1995a,b, 1996) examine the dynamic aspects of the suppression of SOAEs by pulsed external tones, using a Van der Pol oscillator to interpret the results.

The Van der Pol oscillator

A classic example of a limit cycle oscillator is Van der Pol's equation with the general unforced form given by

$$\ddot{x} + (\gamma x^2 - \mu)\dot{x} + \omega_0^2 x = f, \tag{22}$$

where x is the displacement, f is the normalized driving force, and ω_0 is the oscillator natural frequency. The nonlinear damping is given by $\zeta(x) = \gamma x^2 - \mu$, and the restoring force is linear with $k(x) = \omega_0^2 x$. A Hopf bifurcation exists with respect to the parameter μ. If $\mu < 0$, the oscillator displays a simple stable spiral solution. If $\mu > 0$, it is possible to show that the Van der Pol equation has a limit cycle solution via Liénard's theorem, provided that γ is also positive.

For $\mu > 0$, the Van der Pol equation appears as a simple harmonic oscillator with a nonlinear damping term, $(\gamma x^2 - \mu)$. This term acts like positive damping (sink) for large $|x|$ and like negative damping (source) for small $|x|$. It causes large amplitude oscillations to decay, but it pumps energy into them if they become too small. As a result the system reaches equilibrium via a limit cycle that is independent of the initial conditions.

The harmonically forced Van der Pol equation is given by

$$\ddot{x} + (\gamma x^2 - \mu)\dot{x} + \omega_0^2 x = \sqrt{2\beta}\cos \omega t, \tag{23}$$

where ω_0 is the natural frequency of the unforced oscillator, ω is the driving frequency, and β is the rms driving level. The high amplitude response of this oscillator will produce a 0.33 dB/dB slope on the input-output level curve (Camalet *et al.*, 2000). In order to produce similar level curves to those already described in this paper, the high level regions need a 0.5 dB/dB slope. Elliott and Harte (2003) achieved this by altering the classical forced Van der Pol equation to

$$\ddot{x} + (\gamma |x| - \mu)\dot{x} + \omega_0^2 x = \sqrt{2\beta}\cos \omega t, \tag{24}$$

which has a similar behavior for different values of μ to that described above.

A normalized form of Eq. (24) is sought in order to reduce the number of free parameters and to generalize the nonlinear oscillator. We adopt a new time scale normalized by the natural frequency of the oscillator

$$\tau = \omega_0 t \tag{25}$$

and introduce the new variables

$$x_n = \left(\frac{\gamma}{\mu}\right)^{1/2} x \tag{26}$$

and

$$\eta = \frac{\omega}{\omega_0}, \tag{27}$$

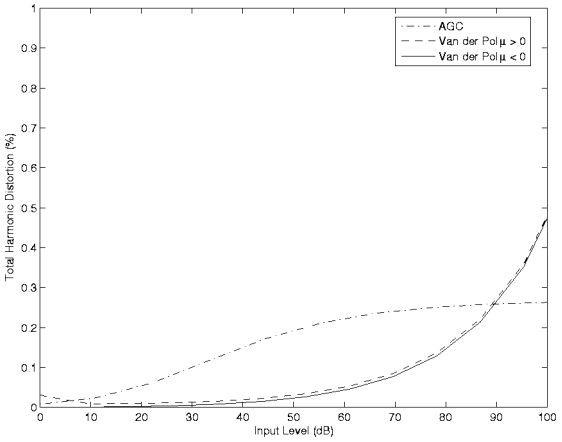


FIG. 6. Total harmonic distortion of the AGC with a normalized time constant $\chi=4$ and the Van der Pol oscillator with both positive and negative low-level damping for a sinusoidal input at the oscillator's natural frequencies.

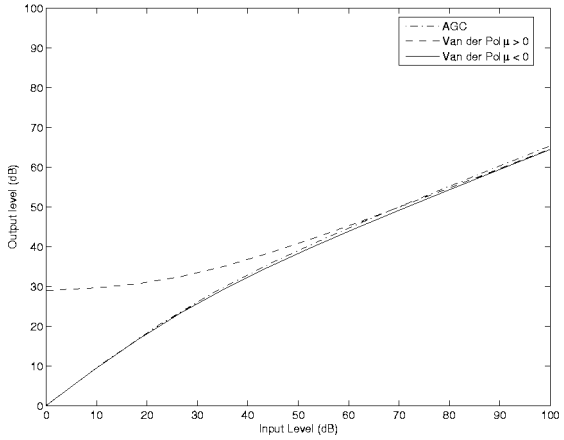


FIG. 7. The level curves for sinusoidal excitation of the AGC (dot-dash) and the Van der Pol oscillator, excited at its natural frequency of oscillation, for $\mu<0$, i.e., passive (solid curve), and $\mu>0$, i.e., active (dashed curve).

where η can be considered a detuning parameter. Obtaining new expressions for the derivative terms, and using normal time-derivative notation without loss of generality, gives rise to the normalized Van der Pol oscillator equation:

$$\ddot{x}_n+\delta(|x_n|-1)\dot{x}_n+x_n=F\cos(\eta\tau), \tag{28}$$

where the parameter

$$\delta=\frac{\mu}{\omega_0} \tag{29}$$

and the normalized forcing term is given by

$$F=\frac{1}{\omega_0^2}\sqrt{\frac{2\beta\gamma}{\mu}}. \tag{30}$$

When numerically simulating the system, in the case where $\mu<0$, so that $\delta<0$, the forcing amplitude becomes complex and thus physically unrealizable. This also affected the stability of the numerical integration routine. It was decided to recast the normalized equation to explicitly take this change into account, so that when $\mu<0$, the normalized Van der Pol equation becomes

$$\ddot{x}_n+\delta(|x_n|+1)\dot{x}_n+x_n=F\cos(\eta\tau), \tag{31}$$

where throughout the various variables definitions μ has been replaced with $|\mu|$.

A state space form of the two variants of the normalized van der Pol oscillator was used in the numerical simulations, and the system was excited by a sinusoid at the natural frequency, ω_0. A Runge-Kutta numerical integration procedure was carried out. The sampling frequency was set to 40 times the natural frequency, with $\delta=0.001$, which affects the level at which the transition is made from linear to nonlinear behavior. Integration was carried out for 10 000 periods of the excitation frequency, to ensure that a steady state had been reached.

The Van der Pol oscillator has the potential to distort the input signal by diverting energy into higher harmonics. Figure 6 shows the numerically estimated total harmonic distortion, $\Xi(\%)$, for both variants of oscillator driven at their natural frequency, plotted as a function of excitation level. It is interesting to note that these curves are almost identical. For both variants, the low level region is dominated by the linear properties of the Van der Pol equations, regardless of whether this linear damping term is acting as a source or a sink. Note that minimal distortion exists for the limit-cycle solution only because the oscillator is being driven at its natural frequency. If this were not the case, then $\Xi(\%)$ would be near 100% at very low levels, since the energy in the system will be at the oscillator natural frequency. As the excitation level increases, the nonlinear damping begins to dominate the dynamics, and thus similar levels of distortion are observed. The highest level of distortion observed for both cases is of the order of 0.5%. Compare with this a static power law nonlinearity with an exponent of $p=0.5$, in Fig. 1, which will produce the same slope on a level curve, but this will result in around 17% distortion in the output waveform. In comparison with an AGC, however, similar levels of distortion are produced if the normalized time constant, χ, is about 4, but in this case the distortion rises from zero at low levels, when the system is approximately linear, to an almost constant level above 30 dB.

C. Input-output level curves

Figure 7 shows the level curves computed from the automatic gain control. The dot-dash curve gives the level curve for an AGC with $\varepsilon=10^3$ and the additional gain a of 10^3, and has a slope of 1 dB/dB at low levels and of 0.5 dB/dB at higher levels.

Figure 7 also shows the level curves for the two Van der Pol oscillators with a low level behavior that is either passive, $\mu<0$, or active, $\mu>0$. The input and output levels were rescaled to correspond to an oscillator with a natural frequency $f_0=1$ kHz, and the results were shown in the figure plotted relative to the level of the $\mu<0$ oscillator excited at 0 dB to aid comparison. The solid line represents the case without the limit cycle solution ($\mu<0$). In the low excitation level region (0–25 dB) a 1-dB increase in excitation produces a corresponding 1-dB increase in the output level. Thus the low level linear region is satisfactorily modeled with this equation. In the high excitation level region (40–100 dB) the level curve has a slope of 0.5 dB/dB. The dashed

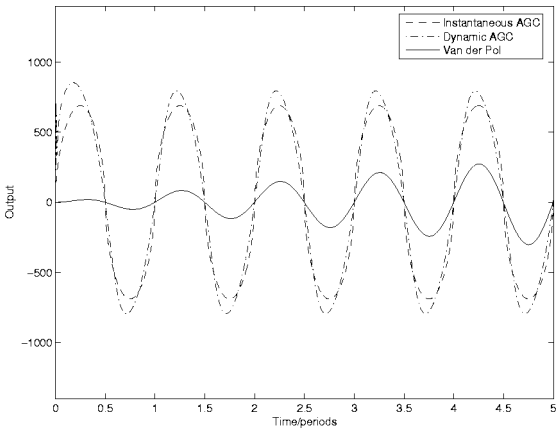


FIG. 8. Output waveform, for the instantaneous AGC system (dashed), AGC with time constant $\chi \approx 0.16$ (dot-dash), and Van der Pol oscillator (solid) all at an input level of 80 dB.

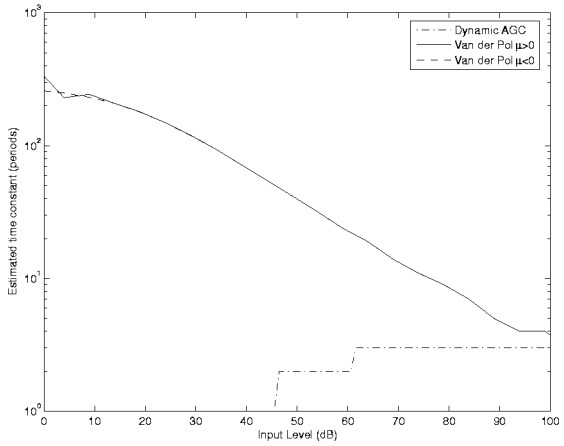


FIG. 9. The time constants of the dynamic AGC ($\chi=4$) and the Van der Pol variants to small perturbations, as a function of level.

line in Fig. 7 represents the oscillator that has a limit-cycle solution. A constant level output is observed for low excitation levels, which then increases for input levels above about 30 dB and asymptotically approaches a 0.5 dB/dB compressive slope at higher levels. The passive Van der Pol thus produces a very similar level curve to the AGC system, but the possibility of negative damping at low input levels can also be used to model SOAEs.

IV. COMPARISONS OF INSTANTANEOUS AND LEVEL-DEPENDENT NONLINEARITIES

Four simple models for cochlear compression have been developed: the input-scheduled power law, the output-scheduled power law (static nonlinear systems), the automatic gain control, and the Van der Pol oscillator (level-dependent nonlinear systems). The output-scheduled power law will be referred to in this section as an instantaneous AGC, as it can be thought of as a limiting case of the AGC system with normalized time constant $\chi=0$. The level curves for the different nonlinear models, as shown in Figs. 3 and 7, are asymptotically very similar. In this section we look at the response of these systems to suddenly applied tonal excitation and for tonal excitation at different levels, in particular examining their transient response and instantaneous input-output characteristic. Although significant differences in these properties are observed in the various models, it is much more difficult to relate these theoretical predictions to physiological measurements. This is mainly because of the need for a fully distributed model to represent transient effects that involve the dispersive properties of the cochlea as well as its nonlinear properties.

A. Transient response

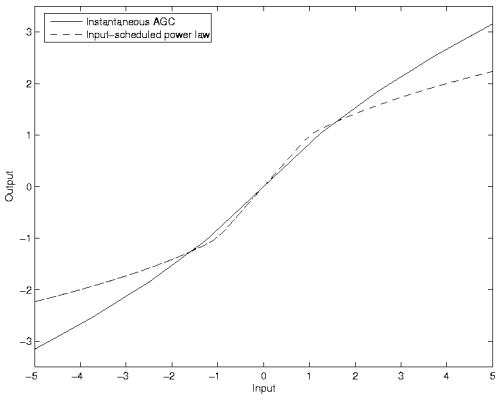
Figure 8 shows the transient response, to a suddenly applied tone, of the instantaneous AGC and the level-dependent system consisting of a feedback AGC with a normalized time constant $\chi \approx 0.16$. Although the level-dependent system takes some time to respond to the input, this time is much less than a cycle of the input waveform for such a well-adjusted AGC. An AGC circuit with a not unreasonable compromise between distortion and transient response has a normalized time constant $\chi \approx 1$. This implies that for a 1-kHz BM site the time constant is of the order of 1 ms.

The transient response of the Van der Pol oscillator at an input level of 80 dB is also shown in Fig. 8. It should be noted, however, that the time constant of the Van der Pol oscillator is very level dependent. Figure 9 shows the time constants associated with the dynamic AGC and the Van der Pol oscillators, defined as the time until the output value has dropped from its initial height to within 36% of the new steady-state value, observed as a function of input level. The time constants for the dynamic AGC are essentially zero (i.e., instantaneous) for low levels of excitation. In this region the AGC is acting linearly and the detector dynamics do not influence the instantaneous gain. As excitation level is raised, the time constant increases to a limiting value. On the other hand, both variants of the Van der Pol equations yield time constants that reduce for increasing excitation level. The higher level region produces time constants about an order of magnitude larger than those found for the well-adjusted AGC system.

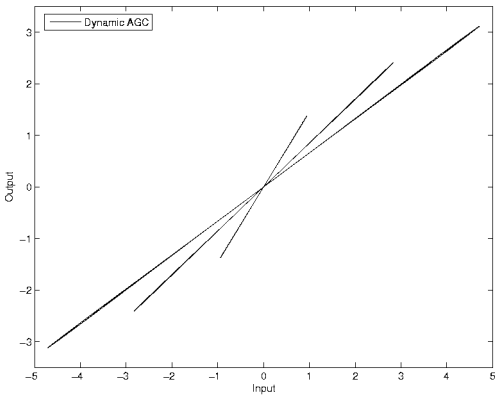
B. Instantaneous input-output characteristic

Figures 10(a)–(c) show the instantaneous value of the output waveform plotted against the instantaneous value of the input waveform, for a tonal excitation of several of the nonlinear models when driven at three different amplitudes. These three input-output characteristics overlay one another for the static nonlinearities [Fig. 10(a)], and thus form a single-valued nonlinear input-output characteristic. The instantaneous input-output characteristics of the AGC with a long time constant [Fig. 10(b)], however, are straight lines, although the slopes of these lines, reflecting the gain of the AGC, change with the amplitude of the input signal.

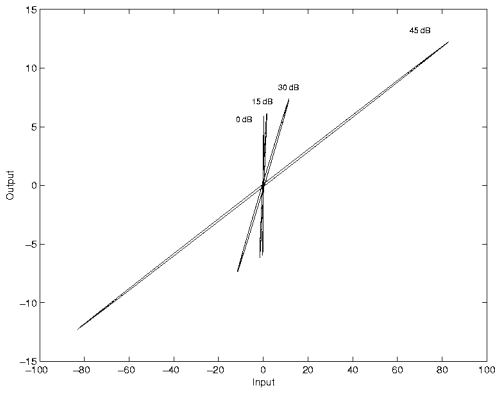
Figure 10(c) shows the plots of the instantaneous input-output characteristics of the Van der Pol oscillator at three excitation levels, for the limit-cycle oscillator with $\mu>0$. The output at low levels is a limit cycle of fixed amplitude, therefore the variation in the y scale is not as large as expected. Regardless of the excitation level in this region, a fixed output is observed which implies a very high ''gain'' at these low levels. The effect of reducing the input level can be



(a)



(b)



(c)

FIG. 10. Instantaneous input-output characteristic for sinusoidal excitation at three levels of the instantaneous nonlinear system, with input-scheduled power law and the instantaneously acting AGC (a), the level-dependent nonlinear system, feedback AGC (b), and the Van der Pol (c).

thought of as the system ''gain'' ramping up to compensate. The characteristics also appear to be straight lines or thin ellipses.[1] This is the case because the linear region of the system dynamics is being excited, and therefore little distortion occurs. Figure 10(c) also shows the effect of higher level excitation. The variation in the output or *y* axis is of a more reduced order of magnitude than for the input or *x* axis, hence the observed change in slope or ''gain'' in this region. This is consistent with the fact that in the region of excitation

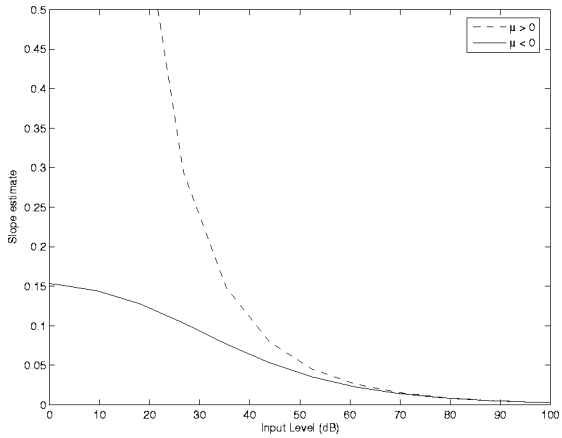


FIG. 11. Instantaneous input-output characteristic slope estimates plotted as a function of excitation level, and each curve normalized for the value of the slope at 0 dB excitation for both $\mu>0$ and $\mu<0$.

levels the oscillator is nonlinear with a compressive level curve.

The low level instantaneous input-output characteristics, of the oscillator with $\mu<0$, show a linear reduction of the output level as the input level is reduced. This results in characteristics with constant slope regardless of the excitation level, thus producing a very different characteristic form than when $\mu>0$. Figure 11 shows an estimate of the slope of the instantaneous input-output characteristic plotted as a function of excitation level. For the limit-cycle case ($\mu>0$) the slope or gain decays rapidly as the excitation level is increased, with the major transition between high and low gain occurring at the knee point on the level curve as expected. For $\mu<0$ the low level region produces a constant gain, due to the system being dominated by linear dynamics, which reduces as the level is increased. The two curves become similar for high level excitation.

Note that the instantaneous input-output characteristics of both the level-dependent AGC and the Van der Pol oscillator demonstrate a multi-valued nonlinearity.

V. DISCUSSION

Several different types of nonlinear system have been investigated as potential models for the inherent nonlinearity within the cochlea. It has been shown that all of these systems can produce input-output level curves that are similar to those measured on the BM, having a 1 dB/dB slope at low levels and 0.5 dB/dB slope, for example, at high levels. The AGC could be made to have a transient response that appears almost instantaneous for tonal excitation by careful selection of the low-pass filter cutoff in the output level detector circuit. The amount of harmonic distortion produced by a dynamic AGC can be far below that produced by a static nonlinearity, however, although both display a similar variation with input level, with little distortion at low levels rising to a constant level above about 30 dB. The limit-cycle oscillator at high excitation levels displays time constants an order of magnitude greater than for the well-adjusted AGC, and, although it displays comparable amounts of distortion, the distortion continues to rise with level, instead of reaching a limiting value.

Another difference between the static and dynamic models, however, is their instantaneous input-output characteristics. For the static nonlinear systems the characteristic is single valued. Different regions of the characteristic are exercised for different excitation levels, but they all lie on the same curve. The dynamic, level-dependent models, however, produce roughly a linear/straight line characteristic, which varies with excitation level. This produces a multi-valued instantaneous input-output characteristic. If the nonlinearity is inherently multi-valued, it is not possible to model or describe such a system with a Volterra, Wiener, or functional series representation. It is not clear, however, how these distortion, transient, and instantaneous characteristics will be affected if these nonlinearities are incorporated into a distributed model of the cochlear, and thus no attempt is made at this stage to compare these predictions with measurements. However, we do note that very different nonlinear models can produce plausible level curves, but have very different behavior in other respects.

There is a well-established history in hearing research of using Volterra and Wiener series to represent the inherent nonlinearities in the auditory pathway (van Dijk *et al.*, 1994, 1997a,b; Eggermont, 1993; Gifford and Tomlinson, 1989; Johnson, 1980; Lasky *et al.*, 1995; Maat *et al.*, 2000; Shi and Hecox, 1991; Wickesberg *et al.*, 1984). Since the structure of the nonlinearity in the cochlea is not fully understood, a parametric identification approach such as these functional representations are often used. It would appear, in light of the findings in this work, that the foundations of these nonlinear system identification approaches may need to be examined in more detail if level-dependent models do emerge as likely contenders, as these functional methods are unable to describe multi-valued nonlinearities. There is already some physical evidence to suggest that the compressive nonlinearity inherent in the cochlea is level dependent. In particular Maat *et al.* (2000) used the polynomial correlation method to access higher-order nonlinearities in noise-evoked otoacoustic emission (NEOAE) data and found that for a given level of excitation the polynomial correlation function above the linear, or first-order, response contributed no significant components but that the linear response depends on the input level. This implies that for a given level of excitation a linear system is sufficient to describe NEOAEs, which would suggest a level-dependent nonlinearity over a static or memoryless characteristic-based system.

If a functional model were to be erroneously applied to a multi-valued nonlinear system, the resulting kernel would only be valid for a particular level of excitation. The derived model then could only describe the response near or on a single operating level.

ACKNOWLEDGMENTS

This work was jointly funded under the EPSRC doctoral training scheme and a Marie Curie Research Fellowship via the European Doctorate in Sound and Vibration Studies. We also gratefully acknowledge Søren Laugesen for his correspondence regarding detector circuit dynamics and level compression from automatic gain controls, as well as Ben Lineton and Steve Billings for informative discussions. We would also like to thank the reviewers for their helpful comments.

[1]It should be noted that the curves in these figures are not the true input-output characteristics; the output has been shifted by $\pi/2$ in phase in order to produce near straight line characteristics. This was done simply to aid visualization of the varying gain concept.

Allen, J. B., and Neely, S. T. (**1992**). ''Micromechanical models of the cochlea,'' Part. Sci. Technol. **45**, 40–47.

Bialek, W., and Wit, H. P. (**1984**). ''Quantum limits to oscillator stability: theory and experiments on acoustic emissions from the human ear,'' Phys. Lett. A **104**, 174–178.

Bian, L., Chertoff, M. E., and Miller, E. (**2002**). ''Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **112**, 198–210.

Bian, L., Linhardt, E. E., and Chertoff, M. E. (**2004**). ''Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **115**, 2159–2172.

Camalet, S., Duke, T., Julicher, F., and Prost, J. (**2000**). ''Auditory sensitivity provided by self-tuned critical oscillations of hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 3183–3188.

Choe, Y., Magnasco, M. O., and Hudspeth, A. J. (**2000**). ''A model for amplification of hair-bundle motion by cyclical binding of ca^{2+} to mechanoelectircal-transduction channels,'' Proc. Natl. Acad. Sci. U.S.A. **95**, 15321–15326.

de Boer, E. (**1996**). ''Mechanics of the cochlea: modelling efforts,'' in *The Cochlea*, 3rd ed., edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). ''Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears,'' J. Acoust. Soc. Am. **110**, 3119–3131.

Eggermont, J. J. (**1993**). ''Wiener and Volterra analysis applied to the auditory system,'' Hear. Res. **66**, 177–201.

Eguíluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A. J., and Magnasco, M. O. (**2000**). ''Essential nonlinearities in hearing,'' Phys. Rev. Lett. **84**, 5232–5235.

Elliott, S. J., and Harte, J. M. (**2003**). ''Models for compressive nonlinearities in the cochlea,'' Institute of Sound and Vibration Research technical memorandum no. 913, May.

Gifford, S. J., and Tomlinson, G. R. (**1989**). ''Recent advances in the application of functional series to non-Linear structures,'' J. Sound Vib. **135**, 289–317.

Gold, T. (**1948**). ''Hearing II: The physical basis of the action of the cochlea,'' Proc. R. Soc. London, Ser. B **135**, 492–498.

Harte, J. M. (**2004**). ''Nonlinear identification for otoacoustic emissions,'' Ph.D. thesis, University of Southampton, Institute of Sound and Vibration Research.

Johnson, D. H. (**1980**). ''Applicability of white-noise nonlinear system analysis to the peripheral auditory system,'' J. Acoust. Soc. Am. **68**, 876–884.

Jülicher, F., and Prost, J. (**1997**). ''Spontaneous oscillations of collective molecular motors,'' Phys. Rev. Lett. **78**, 4510–4513.

Kanis, L. J., and de Boer, E. (**1993**). ''Self-suppression in a locally active nonlinear model of the cochlear: A quasilinear approach,'' J. Acoust. Soc. Am. **94**, 3199–3206.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Lasky, R. E., Maier, M. M., Snodgrass, E. B., Laughlin, N. K., and Hecox, K. E. (**1995**). ''Auditory evoked brainstem and middle latency responses in *Macaca mulatta* and humans,'' Hear. Res. **89**, 212–225.

Long, G. R., Tubis, A., and Jones, K. L. (**1991**). ''Modelling synchronisation and suppression of spontaneous otoacoustic emissions using Van der Pol oscillators: effects of aspirin administration,'' J. Acoust. Soc. Am. **89**, 1201–1212.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3117.

Lyon, R. F. (**1990**). ''Automatic gain control in cochlear mechanics,'' Proceedings of the Mechanics and Biophysics of Hearing Conference, Wisconsin, pp. 395–402.

Lyon, R. F., and Mead, C. (**1988**). ''An analogue electronic cochlea,'' IEEE Trans. Acoust., Speech, Signal Process. **36**, 1119–1133.
Maat, B., Wit, H. P., and van Dijk, P. (**2000**). ''Noise-evoked otoacoustic emissions in humans,'' J. Acoust. Soc. Am. **108**, 2272–2280.
Martin, P., and Hudspeth, A. J. (**2001**). ''Compressive nonlinearity in the hair bundle's active response to mechanical stimulation,'' Proc. Natl. Acad. Sci. U.S.A. **98**, 14386–14391.
Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). ''A computational algorithm for computing nonlinear auditory frequency selectivity,'' J. Acoust. Soc. Am. **109**, 2852–2861.
Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). ''Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**, 2761–2778.
Murphy, W. J., Tubis, A., Talmadge, C. L., and Long, G. R. (**1995a**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. I. Response to pulsed single-tone suppressors,'' J. Acoust. Soc. Am. **97**, 3702–3710.
Murphy, W. J., Tubis, A., Talmadge, C. L., and Long, G. R. (**1995b**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. II. Suppression of interacting emissions,'' J. Acoust. Soc. Am. **97**, 3711–3720.
Murphy, W. J., Tubis, A., Talmadge, C. L., Long, G. R., and Krieg, E. F. (**1996**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. III. Response to a single tone at multiple suppression levels,'' J. Acoust. Soc. Am. **100**, 3979–3982.
Norton, S. J., and Neely, S. T. (**1987**). ''Tone-evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **81**, 1860–1872.
Nuttall, A. L., and Dolan, D. F. (**1996**). ''Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig,'' J. Acoust. Soc. Am. **99**, 1556–1565.
Ospeck, M., Eguíluz, V. M., and Magnasco, M. O. (**2001**). ''Evidence of a Hopf bifurcation in frog hair cells,'' Biophys. J. **80**, 2597–2607.
Pfeiffer, J. O. (**1970**). ''A model for two-tone inhibition of single cochlear-nerve fibres,'' J. Acoust. Soc. Am. **48**, 1373–1378.
Prieve, B. A., Gorga, M. P., and Neely, S. T. (**1996**). ''Click- and tone-burst-evoked otoa7coustic emissions in normal-hearing and hearing-impaired ears,'' J. Acoust. Soc. Am. **99**, 3077–3086.
Rabinowitz, W. M., and Widin, G. P. (**1984**). ''Interactions of spontaneous otoacoustic emissions and external sounds,'' J. Acoust. Soc. Am. **76**, 1713–1720.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mo̊sbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Recio, A. (**2000**). ''Study of mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**, 3317–3332.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar membrane responses to tones at the base of the chinchilla cochlea.,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Sellick, P. M., Patuzzi, R., and Johnston, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Shi, Y., and Hecox, K. E. (**1991**). ''Nonlinear system identification by *m*-pulse sequences: application to brainstem auditory evoked responses,'' IEEE Trans. Biomed. Eng. **38**, 834–845.
Talmadge, C. L., Tubis, A., Wit, H. P., and Long, G. R. (**1991**). ''Are spontaneous otoacoustic emissions generated by self-sustained cochlear oscillators?'' J. Acoust. Soc. Am. **89**, 2391–2399.
van Dijk, P., and Wit, H. P. (**1990**). ''Amplitude and frequency fluctuations of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **88**, 1779–1793.
van Dijk, P., Maat, B., and Wit, H. P. (**1997a**). ''Wiener kernel analysis of a noise-evoked otoacoustic emission,'' Br. J. Audiol. **31**, 473–477.
van Dijk, P., Wit, H. P., and Segenhout, J. M. (**1994**). ''Wiener kernel analysis of inner ear function in the American bullfrog,'' J. Acoust. Soc. Am. **95**, 904–919.
van Dijk, P., Wit, H. P., and Segenhout, J. M. (**1997b**). ''Dissecting the frog inner ear with Gaussian noise. I Application of high-order Wiener-kernel analysis,'' Hear. Res. **14**, 229–242.
Wickesberg, R. E., Dickson, J. W., Morton Gibson, M., and Geisler, C. D. (**1984**). ''Wiener kernel analysis of responses from anteroventral cochlear nucleus neurons,'' Hear. Res. **14**, 155–174.
Yates, G. K. (**1990**). ''Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hear. Res. **50**, 145–162.
Zwicker, E. (**1979**). ''A model describing nonlinearities in hearing by active processes with saturation at 40 dB,'' Biol. Cybern. **35**, 243–250.
Zwicker, E. (**1983**). ''Delayed evoked oto-acoustic emissions and their suppression by Gaussian-shaped pressure impulses,'' Hear. Res. **11**, 359–371.
Zwicker, E. (**1986**). ''A hardware cochlear nonlinear preprocessing model with active feedback,'' J. Acoust. Soc. Am. **80**, 146–153.
Zwicker, E., and Schloth, E. (**1984**). ''Interrelation of different otoacoustic emissions,'' J. Acoust. Soc. Am. **75**, 1148–1154.

A computer model of the auditory-nerve response to forward-masking stimuli

Ray Meddis[a] and Lowel P. O'Mard
Centre for the Neural Basis of Hearing at Essex, Department of Psychology, University of Essex, Colchester, CO4 3SQ, United Kingdom

(Received 8 December 2004; revised 28 February 2005; accepted 28 February 2005)

A computer model of the auditory periphery is used to study the involvement of auditory-nerve (AN) adaptation in forward-masking effects. An existing model is shown to simulate published AN recovery functions both qualitatively and quantitatively after appropriate parameter adjustments. It also simulates published data showing only small threshold shifts when a psychophysical forward-masking paradigm is applied to AN responses. The model is extended to simulate a simple but physiologically plausible mechanism for making threshold decisions based on coincidental firing of a number of AN fibers. When this is used, much larger threshold shifts are observed of a size consistent with published psychophysical observations. The problem of how stimulus-driven firing can be distinguished from spontaneous activity near threshold is also addressed by the same decision mechanism. Overall, the modeling results suggest that poststimulatory reductions in AN activity can make a substantial contribution to the raised thresholds observed in many psychophysical studies of forward masking.
[DOI: 10.1121/1.1893426]

PACS numbers: 43.64.Bt, 43.64.Ld [BLM] Pages: 3787–3798

I. INTRODUCTION

In physiological experiments, the responsiveness of the auditory nerve (AN) is reduced immediately following acoustic stimulation. In psychophysical experiments, thresholds are temporarily raised following similar acoustic stimulation. In both cases, the changes depend on the level and the duration of the preceding stimulation. Recovery, in both cases, occurs over a period lasting more than 100 ms. These superficial similarities suggest that the two phenomena may be linked. Indeed, both physiologists and psychophysicists use the same term, ''forward masking,'' to describe the respective phenomena.[1]

However, Relkin and Turner (1988) have indicated the potential for confusion inherent in an uncritical equation of the two phenomena. They point out that a reduction in physiological response is not the same thing as an increase in perceptual threshold, and that different measurement methods are involved in their respective estimation. When they attempted to apply psychophysical threshold estimation techniques to single-fiber AN observations, they found that the poststimulatory reduction in threshold was considerably less than that seen in human psychophysics. They concluded that the increase in psychophysical thresholds must occur later in the processing system. Specifically, they claim that the nervous system does not use an optimum decision strategy to evaluate the information present in the AN response.

A number of physiological processes have been proposed as possible contributors to psychophysical forward masking. These include peripheral frequency selectivity (Duifhuis, 1973), adaptation in the auditory nerve (Smith, 1977, 1979), efferent inhibitory processes (Shore, 1998), and persistence of neural activity (for example, Plomp, 1964; Moore *et al.*, 1988). See Oxenham (2001) for a brief, recent overview of the issues. It is not intended, in this study, to decide between these alternatives because all are likely to make some contribution. In any case, as Oxenham (2001) has shown, it is difficult to separate their relative contributions experimentally. However, the Relkin and Turner (1988) study and another similarly motivated study (Turner *et al.*, 1994) could create the impression that forward masking does not have its origin in adaptation in the AN. This conclusion goes beyond what is warranted by the data, and it is the purpose of the present study to use computational methods to investigate the matter further.

The study employs an already-published model of the mammalian auditory periphery to simulate AN responses with appropriate poststimulatory adaptation and recovery characteristics. It will be demonstrated that the model is appropriate for the purposes of the investigation by showing that it can be tuned to simulate observations of depression and recovery of the chinchilla AN response following stimulation (Harris and Dallos, 1979). It will also be shown that the model can be used to replicate the findings of Relkin and Turner (1988) referred to above. Finally, a simple coincidence-detection mechanism mimicking the response of a first-order neuron in the cochlear nucleus will be modeled. When this takes as its input the combined response of a small number of AN fibers, it will be demonstrated that substantial increases in threshold do occur. These increases are considerably greater than those seen in Relkin and Turner's study and comparable to those seen in psychophysical studies.

The coincidence-detection mechanism is proposed as a solution to a more general problem in auditory physiology

[a]Author to whom correspondence should be addressed. Current address: Department of Psychology, University of Essex, Colchester, CO4 3SQ, United Kingdom. Electronic mail: rmeddis@essex.ac.uk

concerning how the nervous system distinguishes between spontaneous AN activity and activity driven by acoustic stimulation. While it is natural to believe that the system does this by detecting an increase in the average firing rate of individual auditory-nerve fibers, the physiological mechanism underlying this kind of detection strategy remains problematic. For example, Relkin and Turner's decision-making process implies that the system works by establishing a running count of action potentials (APs) and assesses its significance in association with an expected mean and variance associated with spontaneous firing rate. This optimal decision method is difficult to implement directly in terms of the known anatomy and physiology of the auditory nervous system because it requires secondary mechanisms to estimate and deploy running estimates of these two statistical quantities. This is difficult to describe in terms of the operation of nerve fibers and nerve cells. However, ''coincidence-detection'' is a simple alternative that achieves the same objective.

Spontaneous activity in one AN fiber is uncorrelated with that in other fibers. As a consequence, the occurrence of near-simultaneous APs is relatively rare in a small group of fibers such as might innervate a single cell in the cochlear nucleus. However, when APs are driven by acoustic stimulation, the overall firing rate will increase and the number of near-simultaneous events will also increase. Coincidental firing will also be more probable near the onset of the stimulation because similar AN fibers will have similar first-spike latencies (Heil and Neubauer, 2001). In the case of low-frequency sounds, phase locking to the stimulus will further increase the amount of coincidental activity. On all three counts, coincidental firing is expected to be greater following acoustic stimulation than in silence. If a cell in the cochlear nucleus responds only to coincidental firing of AN fibers, it will respond almost exclusively to acoustic stimulation and only very rarely to spontaneous activity in the AN fibers. This system is consistent with the architecture of the auditory brainstem, where neurons integrate information across a number of input fibers when generating responses.

After a description of the model, this account is divided into three sections, each describing a different evaluation of the model. First, the recovery of the response of the model after stimulation is compared with the data of Harris and Dallos (1979). Second, the model is used to replicate the study of Relkin and Turner (1988). Third, the coincidence detection mechanism is introduced and the model is evaluated using thresholds obtained using a forward masking paradigm (Jesteadt *et al.*, 1982).

II. MODEL DESCRIPTION

The computer model of the mammalian auditory periphery to be used in this evaluation was presented by Sumner *et al.* (2002, 2003a, 2003b) for the guinea pig auditory periphery. Exactly the same model is used except where specifically indicated. This model consists of a chain of separate modules representing, respectively, the pre-emphasis filtering of the outer/middle ear, the vibration of the basilar membrane, the inner hair cell (IHC) receptor potential, IHC/AN synaptic activity, and AN activity. This model is capable of simulating a wide range of physiological measurements but has not yet been evaluated in the context of recovery following stimulation. A full account of the model will not be repeated here and the reader is referred to the descriptions given by Sumner *et al.* Parameters for the model can be found in Sumner *et al.* (2002, Tables I^2 and II; 2003b, Table I). The parameters for high, medium, and low spontaneous rate (HSR, MSR, and LSR) fibers are taken from Sumner *et al.* (2003b, Table I). The original model was tuned using guinea pig data, but the simulations to be described involve using chinchilla and human data. The assumption is made that the underlying processes are the same as far as our present purposes are concerned for all these species.

A. Middle-ear filtering

Middle-ear filtering is modeled by two cascaded linear bandpass Butterworth filters. One filter is second order, with lower and upper cutoffs of 4 and 25 kHz. The second is third-order and has lower and upper cutoffs of 0.7 and 30 kHz. A scaling factor is used to convert sound pressure to realistic stapes velocities (see the changed parameters below).

B. Mechanical filtering

A dual-resonance nonlinear (DRNL) filter is used to simulate the mechanical bandpass filtering of the basilar membrane (BM) in the cochlea. The construction and properties of the DRNL filter have been described in detail elsewhere (Meddis *et al.*, 2001; Sumner *et al.*, 2003b). The DRNL filter consists of two pathways, one linear and the other nonlinear. The signals at the output of the two pathways are combined by simple summation to produce the filter output. Each pathway consists of a cascade of a number of first-order gammatone filters (Patterson *et al.*, 1988) and a cascade of second-order Butterworth low-pass filters.

In the nonlinear pathway, a compression function is located in the middle of the cascade of four gammatone filters between the second and third filter. The linear pathway consists of a cascade of three gammatone filters but does not contain nonlinearity. It has a center frequency that is different from that of the linear path; this contributes asymmetry to the overall filter function. Additional low-pass filters are also employed to improve the fit to animal data. The shape of the bandpass filters is not critical to this study because the stimuli to be used are all at the same frequency, and the best frequency of the filter was always set at the frequency of the stimulus. However, the nonlinear compression will affect the pattern of growth of forward masking at high masker levels.

C. Inner hair cell (IHC) receptor potential

The process of mechanical to electrical transduction is simulated by a model of the IHC receptor potential that converts BM velocity to IHC receptor potential through the deflection of the IHC stereocilia.

The receptor potential controls the rate of calcium influx into the cell at the synapse. The calcium is accumulated in a leaky integrator, and the instantaneous level of calcium controls the rate of transmitter release. In the model, the differ-

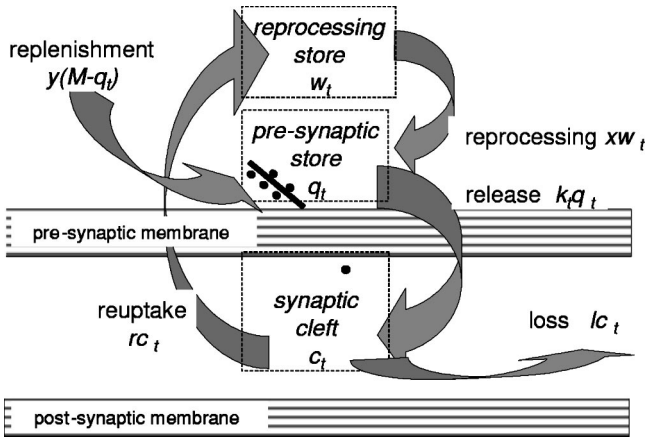


FIG. 1. Flow of transmitter in a model of the Inner hair cell/auditory-nerve synapse. See the text for full description.

ences between fiber types (HSR, MSR, and LSR) are the result of different rates of influx of calcium into the presynaptic sites. HSR sites have a faster rate of influx and correspondingly higher spontaneous release rates. The rate of calcium influx into the IHC for a given receptor potential is specified by two parameters: the maximum rate of influx, $GCa_{\max}$, and a threshold parameter, GCa_{thr}. These parameters are unchanged from Sumner *et al.* (2003b, Table I).

D. IHC synapse

In this study, the IHC/AN synapse is of particular interest. The depression of response following a masker is assumed to result from a depletion of available transmitter substance (Smith, 1977, 1979). Figure 1 illustrates how the model handles the flow of transmitter. A small number, $q(t)$, of vesicles of transmitter is held in a presynaptic store. Each vesicle has an equal probability, $k(t)d(t)$, of being released in a single epoch of duration dt. $k(t)$ is a function of the accumulated presynaptic calcium.

When transmitter is released into the cleft, the amount of presynaptic transmitter, $q(t)$, necessarily falls and the likelihood of further release events is reduced proportionately; the model is said to have ''adapted.'' Transmitter material is recycled as follows. Reuptake into the cell occurs at a rate $rc(t)$, where $c(t)$ is the amount of transmitter in the synaptic cleft and r is the rate of reuptake of transmitter into the cell. Some transmitter is lost from the cleft at a rate $lc(t)$, where l is the rate of loss of transmitter from the cleft. After reuptake, transmitter material is reprocessed and returned to the available presynaptic pool at a rate $xw(t)$, where $w(t)$ is the amount of material in the reprocessing store and x is the rate at which transmitter is returned from the reprocessing store to the pre-synaptic available store. The loss of transmitter material from the cleft is slowly compensated by a replenishment mechanism that supplies new transmitter at a rate $y[M\text{-}q(t)]$, where M is maximum capacity of the available store and y is the rate at which new transmitter is made available. Transfer of vesicles in and out of the presynaptic store is a quantal stochastic process $N(n,\rho)$, where each of n possible events has an equal probability, ρdt, of occurring in a single simulation epoch. The flow of transmitter is specified by the following differential equations:

$$dq(t)/dt = N[w(t),x] + N([M\text{-}q(t)],y) - N[q(t),k(t)], \quad (1)$$

$$dc(t)/dt = N[q(t),k(t)] - lc(t) - rc(t), \quad (2)$$

$$dw(t)/dt = rc(t) - N[w(t),x]. \quad (3)$$

After prolonged and intense stimulation, the available store can become completely depleted. If forward masking is partly explained by depletion of neurotransmitter at the IHC/AN synapse, the time course of recovery from forward masking will be partly governed by the rate at which the store of available transmitter is refilled. Recovery from depletion takes place only through the reprocessing and replenishment routes. As a consequence, parameters x and y are important determinants of the rate of recovery. These two parameters of the model will be adjusted below to match recovery measurements made in the AN.

E. AN response

An action potential is deemed to occur when one or more vesicles are released into the synaptic cleft so long as the postsynaptic AN fiber is not already in a refractory state. An absolute refractory state follows an action potential and lasts for 0.75 ms and the exponential recovery of responsiveness following the relative refractory period has a time constant of 0.6 ms.

The response of the fiber can be evaluated either as a probability of firing (probabilistic mode) or by generating individual spike events using a random number generator (stochastic mode). The former is much quicker to compute and will be used for exploring AN recovery from adaptation (Sec. III). However, for threshold measurements (Secs. IV and V), the model is evaluated in stochastic mode.

F. Model implementation

The model was evaluated at an update rate of 100 kHz. It was implemented using the Development System for Auditory Modelling (DSAM), a library of computer modules simulating each of the stages of the auditory model described above. The software used in this study is available on the Worldwide Web at http://www.essex.ac.uk/psychology/hearinglab/

G. Parameter changes

The scaling factor in the middle-ear simulation was changed from its previous value of $1.4E-10$ to $6E-10$ m/s/μPa. This increases the middle-ear gain by 12.6 dB relative to the previous value and gives AN rate-level functions with similar thresholds to those given in Harris and Dallos (see Fig. 2).

Two IHC/AN synapse parameters were altered to give an improved fit to the chinchilla forward-masking data reported by Harris and Dallos (1972). The rate of transmitter replenishment, y, was reduced from 10 to 3 vesicles/s and the reprocessing rate, x, was reduced from 66 to 30 vesicles/s.

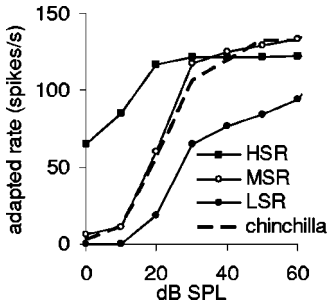


FIG. 2. Rate/level function for model HSR, MSR, and LSR fibers. The dashed line shows a chinchilla rate-level function taken from Harris and Dallos (1979, Fig. 9). Their function is normalized to the rate response of the model MSR fiber at 60 dB SPL.

These changes had little effect on the previously published properties of the model except for a slowing in the time constant of short-term adaptation from 80 to 120 ms. This is longer than that reported for the gerbil by Westerman (1985), but consistent with reports by Smith (1979) in gerbil and Yates *et al.* (1985) in guinea pig. Chimento and Schreiner (1990) report adaptation time constants of 125 ($\pm$52) ms in cat. The model short-term adaptation was measured starting 20 ms after the stimulus onset. Yates *et al.* (1985) claim to have seen an association between the rate of adaptation and the rate of recovery in their observations. This would be consistent with the change observed in the model. Unfortunately, they do not supply any quantitative description of this effect.

III. EVALUATION: AN RECOVERY FROM ADAPTATION

The model was evaluated using a forward-masking paradigm and the same stimuli as those described in Harris and Dallos (1979). Except where stated, the stimulus characteristics were as follows. All maskers and probes were pure tones presented at BF (5750 Hz). Stimulus durations were 100 ms for maskers and 15 ms for probes. The gaps between masker and probe were 1, 2, 5, 10, 20, 50, 100, and 300 ms. Both masker and probe had a rise and fall time of 1 ms using a raised-cosine ramp function. The response to the probe was measured as the total number of spikes observed during the presentation of the probe after allowing for the latency of the response.

A. Rate-level functions

Model rate-level functions for BF tones of 6 kHz are given in Fig. 2. The rate-level functions show adapted rates based on the last 50 ms of the response to the 100-ms masker tones. Functions were measured for three fiber types, HSR, MSR, and LSR. The rate-level functions for the MSR fiber were used to establish the best setting for the stapes gain parameter (see Sec. II G) using a comparison with a single chinchilla fiber given by Harris and Dallos (1979, Fig. 9).

B. Response recovery

Figure 3 illustrates the response of the model to the forward-masking paradigm. The response to the probe increases as the gap between the masker and the probe widens. Figure 4 quantifies this effect for all three types of fiber and

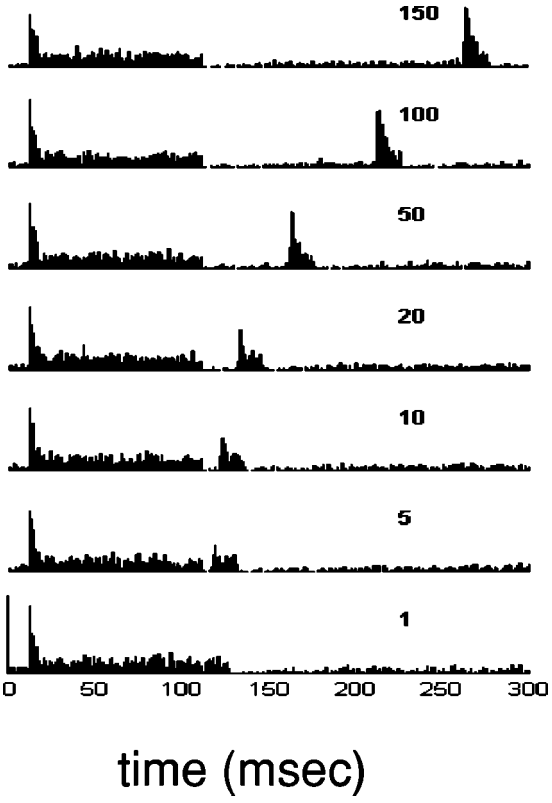


FIG. 3. PSTH response of a model HSR fiber at a range of different masker–probe gaps (ms, indicated in the figure). Masker and probes are all presented at 20 dB above threshold. For this illustration, the model was evaluated in stochastic mode.

a range of masker levels. An example of the data from Harris and Dallos is also given for comparison purposes. Recovery functions based on previously published parameters were found to be too fast (not shown) and the reprocessing and replenishment rates have been adjusted as described to give a more acceptable quantitative fit (see below) to the animal data.

For model HSR fibers, the depression in response at short masker–probe gaps increases over a narrow range of masker levels. For MSR and LSR fibers, this range is considerably greater. Harris and Dallos also found that the depression in responding ceases to grow at some high masker level that is normally the same as the level at which the rate-level function saturates. This possibility was examined in the model. Figure 5 compares the rate-level function for each fiber with the depression of response to the probe immediately after the masker. Both sets of data are normalized between 0 (minimum response) and 1 (maximum response). The functions are closely matched for all three types of fibers.

Longer maskers were shown by Harris and Dallos to give rise to a greater depression in probe response than shorter maskers. Figure 6 summarizes the response of all three model fiber types to maskers of different durations. All three model fiber types show the expected effect.

C. Rate of recovery

At low masker levels, Harris and Dallos found that recovery could be described by a simple exponential recovery

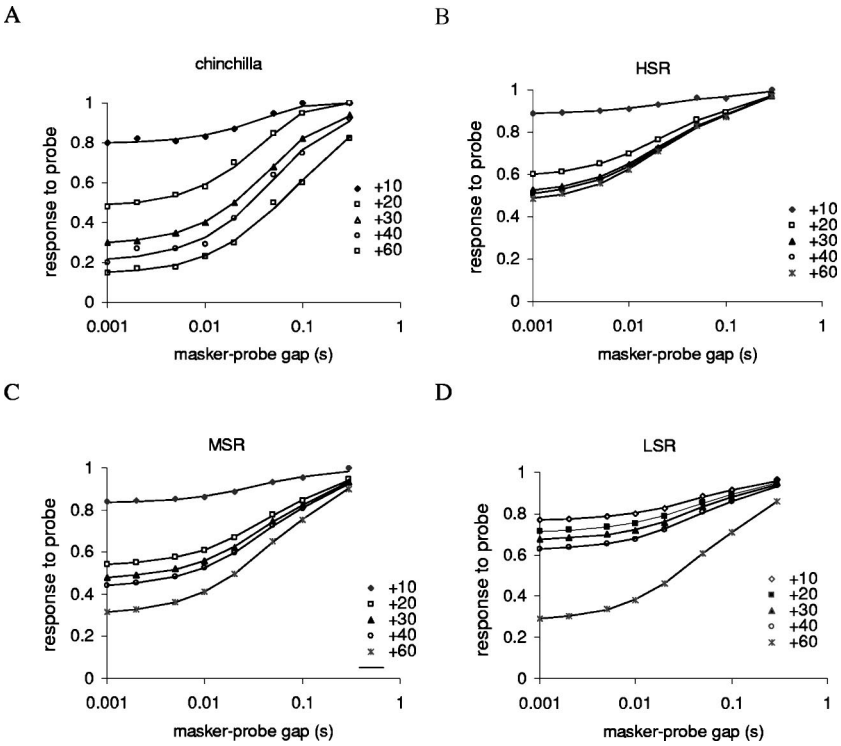


FIG. 4. Response to the probe following stimulation as a function of the masker–probe gap for different masker levels (indicated). All levels are relative to the rate threshold. Response magnitude is the total number of spikes during the presentation of the probe normalized to the probe response measured in the absence of a masker. Continuous lines are double-exponential functions fit to guinea pig and model fiber data. The function fitted to each data set has the form $1-a\exp(-t/\tau_a)-b\exp(-t/\tau_b)$. The parameters used in the fitting process are shown in Table I. (A) Data from Harris and Dallos (1979, Fig. 10, median of 37 fibers). (B)–(D) Model response (HSR, MSR, and LSR fibers, respectively).

function, $1-a\exp(-t/\tau)$, where t is the time since the offset of the masker. However, for more intense masker levels a simple exponential recovery function was a poor fit. Harris and Dallos assumed that the rate of recovery was changing during recovery. Accordingly, they employed different exponential fits to their data for short gaps (<50 ms) and long gaps. Using this approach at higher masker levels, it was necessary to specify short recovery time constants (30 to 70 ms) for the early part of the recovery, but long recovery time constants (30 to 290 ms) towards the end of recovery. Also, longer time constants were required to fit the recovery following higher level maskers.

A reanalysis of their chinchilla data (their Fig. 10) based on the median thresholds of 37 fibers was carried out. It showed that a double exponential recovery process, $1-a\exp(-t/\tau_a)-b\exp(-t/\tau_b)$ gives a good account of their data without the need to change the time constants either as a function of time or masker intensity. The fit to their data is shown in Fig. 4(A) (continuous lines) and the numerical values of the parameters are shown in Table I. The analysis indicates that two time constants (40 and 280 ms) characterize all the data in the figure so long as the coefficients a and b are allowed to vary as a function of masker level. A good fit using a double-exponential recovery process might mean that two separate recovery functions are at work. If so, it makes the animal data qualitatively consistent with the model which has two recovery processes (reprocessing and replenishment).

The model responses are described using the same function [Figs. 4(B)–(D), Table I]. The slow recovery time constant is substantially shorter (158 ms) for model HSR fibers than for both MSR and LSR fibers (228 ms). It is, therefore, a prediction of the model that LSR fibers will recover more slowly given comparable stimulation. Note that all three model fiber types have the same reprocessing and replenishment rate parameters. The between-fiber differences in the observed time constants are a consequence of the different rates of calcium influx into the cells; this is the only difference in the model between fiber types.

Despite the parameter changes, the model recovery is

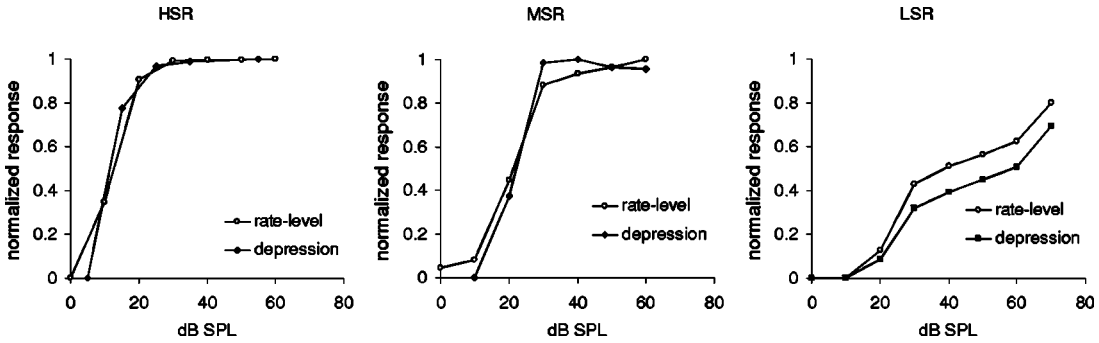


FIG. 5. Comparison of rate-level function and extent of the depression in the response to the probe at the shortest (1-ms) masker–probe gap. All data are normalized between 0 (minimum response) and 1 (maximum response). The LSR did not saturate and the data were normalized to the maximum at 90 dB SPL (not shown).

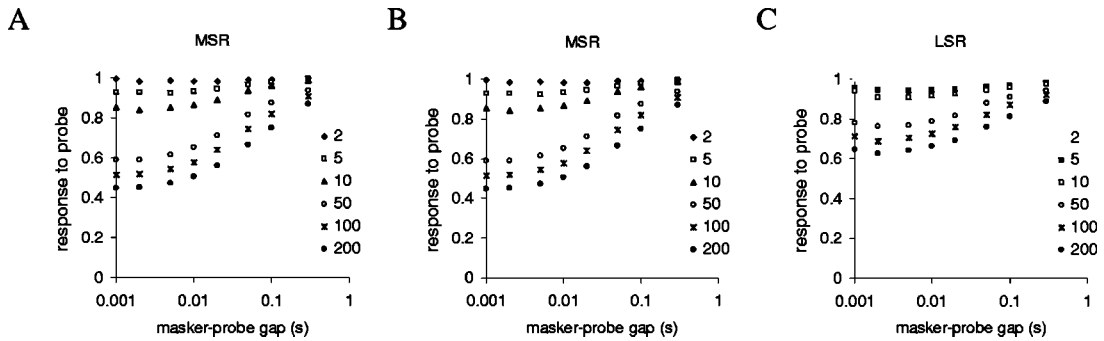


FIG. 6. Normalized recovery functions for HSR, MSR, and LSR fibers as a function of masker *duration*. Both maskers and probes are 20 dB above rate threshold for the individual fiber. Masker durations (ms) are indicated.

still somewhat faster than the median data of Harris and Dallos. Recovery can be slowed further by reducing the values of parameters x and y. However, the slower reprocessing leads to reduced firing rates and a rebalancing of many parameters is required to re-establish typical AN fiber function. The difference between the model performance and the animal data is small enough to justify keeping the other parameters as previously published in the interest of simplicity and transparency. Certainly, the model recovery functions are well within the range of functions illustrated in Harris and Dallos' report, some of which show recovery faster than the model.

D. Recovery of spontaneous rate

Harris and Dallos demonstrated that the time course of recovery from forward masking runs parallel to the recovery of spontaneous activity after stimulation. In the model, both functions are dependent on amounts of transmitter in the available presynaptic store and therefore both are expected to recover at the same rate. Figure 7(A) shows the recovery of spontaneous activity in the HSR model fiber following a masker tone whose level is 20 dB above threshold. Figure 7(B) systematically compares the recovery of spontaneous activity with the recovery of the response to a probe of the same level as the masker. Spontaneous recovery was measured using probe-width windows positioned where the probe would be when using a forward-masking paradigm. The two functions are virtually identical.

TABLE I. Coefficients and time constants used to generate the best-fit functions in Fig. 4. The time constants are constrained to be the *same* for a given data set irrespective of masker level. Coefficients a and b are allowed to vary. They represent the relative contribution of reprocessing and replenishment respectively for a particular masker level.

	Masker level				
	+10	+20	+30	+40	+60
Chinchilla					
a	0.20	0.52	0.52	0.53	0.35
τ_a	**0.04**				
b	0.00	0.00	0.19	0.27	0.51
τ_b	**0.28**				
HSR					
a	0.06	0.23	0.28	0.29	0.31
τ_a	**0.016**				
b	0.06	0.19	0.22	0.22	0.23
τ_b	**0.158**				
MSR					
a	0.11	0.25	0.27	0.29	0.34
τ_a	**0.030**				
b	0.06	0.22	0.26	0.28	0.36
τ_b	**0.227**				
LSR					
a	0.14	0.17	0.19	0.21	0.33
τ_a	**0.035**				
b	0.06	0.09	0.12	0.15	0.37
τ_b	**0.228**				

IV. EVALUATION; AN THRESHOLD INCREASE

Relkin and Turner (1988) estimated threshold changes following stimulation using a measurement paradigm (Relkin and Pelli, 1987) that was chosen to be as similar as possible to the psychophysical two-alternative forced choice (2AFC) methodology normally used in behavioral threshold measurements. Two intervals were used. One contained a masker followed by a probe (''probe'' condition), while the other contained the masker alone (''no-probe'' condition) as shown in Fig. 8(A). They counted the number of spikes observed in single chinchilla AN fibers both to the probe in the probe condition and to a corresponding time window in the no-probe condition. A comparison of the two responses was made and the level of the probe adjusted on the basis of the result of the comparison. The interval containing the greater number of spikes was chosen as the interval deemed to contain the probe. The decision to increase or decrease probe level in the adaptive procedure was based on four trials. If the total number of correct detections of the probe was three or more, the level was decreased. Otherwise it was increased. Using this procedure, the average probe level converges on the level for which the probability of a correct detection is 0.66.

This approach allows a direct comparison between threshold shifts obtained using AN response and those observed in psychophysical experiments. They found that AN threshold shifts were substantially smaller than behavioral threshold shifts. For HSR fibers, threshold shifts using the most intense maskers were in the region of 3–20 dB, while for LSR fibers the shifts were between 8 and 21 dB. These are considerably smaller than the maximum shifts observed in human listeners using similar stimuli, sometimes as large as 70 dB (Widin and Viemeister, 1979).

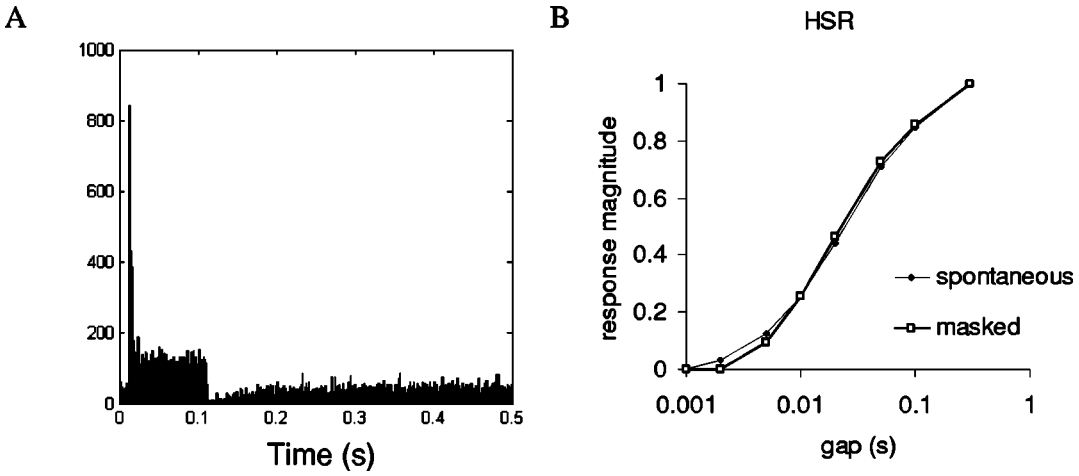


FIG. 7. Relationship between recovery of spontaneous rate and recovery from forward masking in the model HSR fiber. All rates are normalized to have zero minimum and unit maximum. (A) PSTH of a model HSR fiber in response to a 100-ms masker presented at 20 dB above rate threshold evaluated in stochastic mode. (B) Recovery of spontaneous rate following the same masker (no probe) compared with the recovery of the probe response (evaluated in probabilistic mode).

Their experimental paradigm was replicated as closely as possible using the computer model of the auditory periphery described above. No changes were made to any of the parameters of the model. The masker was a 102-ms BF tone and the probe was a 25-ms BF tone. Both tone durations included 2-ms cosine-squared onset and offset ramps and were presented at 5 kHz. The total duration of each interval was 0.4 s. There was no gap between the masker and the probe. Thresholds were estimated for masker levels between −10 and 80 dB SPL in steps of 5 dB. The probe stimulus was initially set to 50 dB SPL and changed in steps of 5 dB for six reversals. The probe was then adjusted in steps of 0.5 dB for 14 more reversals. On successive trials the level of the probe stimulus was increased or decreased according to the principles described above. Threshold was then estimated as the mean of the last 12 reversals. Thresholds were averaged over five blocks of trials. The number of APs recorded during the probe presentation was typically small (<5) and ties often occurred between probe and no-probe conditions. When this happened a random choice was made between the two intervals.

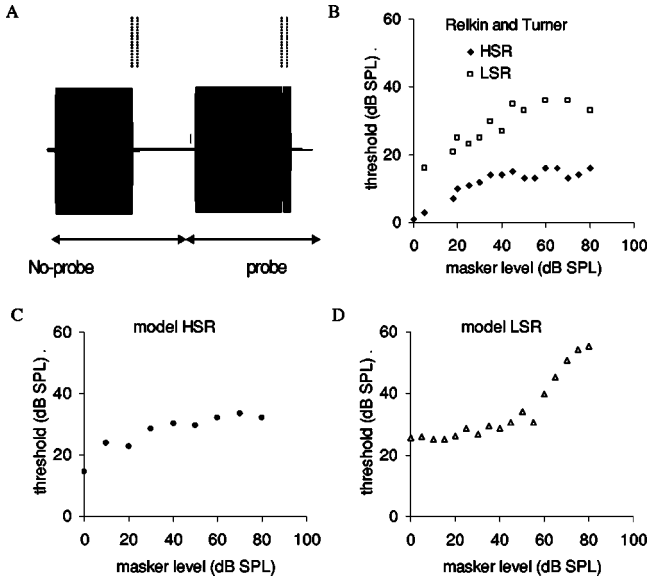


FIG. 8. Simulation of an experiment of Relkin and Turner (1988, Fig. 5). (A) Stimuli were presented in two intervals. Both intervals contained a masker but only the second interval contained a probe. The observation windows used for counting spikes in both intervals are shown as vertical dotted lines. (B) Chinchilla thresholds for two example fibers as a function of masker level; redrawn from Relkin and Turner (1988). (C), (D) Model thresholds as a function of masker levels for an HSR and an LSR fiber. Each threshold estimate is the mean of five trials.

Thresholds as a function of masker level are shown in Figs. 8(C) and (D). Probe thresholds increase with masker level. The model HSR fiber showed an 18-dB shift, while the results of Relkin and Turner [Fig. 8(B)] for two selected fibers indicate a smaller expected threshold shift of about 12 dB for a fiber with a spontaneous rate of 60 spikes/s (their Fig. 6). The discrepancy is well within the variation among animal fibers.

The model LSR fiber [Fig. 8(D)] shows a maximum 25-dB threshold shift, which is comparable with the 20-dB shift for the chinchilla LSR fiber. However, the rise in the model function is shifted to the right relative to the chinchilla data [Fig. 8(B), white squares]. The model LSR shows no threshold shift below 20 dB SPL and does not saturate at high masker levels. This is what we would expect of a typical LSR fiber. Relkin and Turner's data, on the other hand, show unexpected increases in masking for low-level maskers (<20 dB SPL). They claim in their report that probe thresholds (expressed as a function of masker level) closely follow the rate-level function in most of their fibers. Normally LSR fibers have high thresholds and the rise in masking should, therefore, occur only at high masker levels (>30 dB SPL). Unfortunately, the corresponding chinchilla rate-level function is not illustrated in their report. If the animal LSR function is shifted to the right by 30 dB, the discrepancy between the animal and model data would largely disappear. For this reason, the difference is not regarded as serious.

The response of the model is consistent with Relkin and Turner's observations in two important respects. The pattern of threshold shifts across masker levels is consistent with a fiber's rate-level function and the overall shifts in threshold

are too small to serve as an explanation of the large shifts routinely observed in human psychophysical studies.

V. EVALUATION; PSYCHOPHYSICAL FORWARD MASKING

A. Detection mechanism

The aim of the study is to determine whether AN recovery following stimulation as observed in animals can help explain psychophysical forward-masking data in humans. Relkin and Turner's results indicate that an optimal processor attached to a single AN fiber would produce only small threshold shifts in a psychophysical forward-masking experiment. These shifts are too small to explain the large shifts observed in studies with human listeners. Relkin and Turner suggest that some kind of suboptimal processing must follow the auditory nerve. Of course, this could occur at any later stage in the auditory nervous system. The evaluation to be described considers the possibility that the large threshold shifts occur at the first auditory synapse after the AN.

The detection method relates to the general question of what kind of post-AN processing might be involved in the detection of an acoustic signal near threshold. At present there is no generally accepted answer to this important question. One common view is that signals are detected when the firing rate of an AN fiber increases above spontaneous activity. While this may be true, it does not say how that decision is made in terms of physiological structures. The solution proposed below is that small groups of AN fibers converge on single units in the cochlear nucleus (CN) and a ''hit'' decision is made when a number of these respond nearly simultaneously and cause the CN unit to respond with an action potential of its own. This requirement of near-simultaneity is not the same as the spike count across the full stimulus duration as measured by Relkin and Turner. Only spikes that occur very close together in time influence the output of the detector; other spikes are ignored.

The requirement of near-simultaneous firing has the useful consequence of making the system relatively insensitive to spontaneous activity. Spontaneous activity is uncorrelated across fibers, and the probability of coincidental firing across, say, three or more fibers is very low, especially if the coincidence window is narrow. The proposal also meets the ''suboptimal'' requirement of Relkin and Turner in the sense that not all the information present in the AN response is used in the decision; only near-simultaneous spikes influence the decision. A third attraction of this approach is that it is physiologically plausible. AN fibers converge on cells in the cochlear nucleus and all CN cell types receive inputs from a number of fibers. It is highly likely that some of these cells require near-simultaneous AP inputs from a number of fibers before the cell responds with its own AP (Ferragamo *et al.*, 1998 in multipolar cells; Golding *et al.*, 1995 in octopus cells).

The model to be evaluated specifies a fixed number of AN fibers converging on a single CN unit that only responds when a minimum number of AN spikes occur within a window of specified width. For example, a group of 10 AN units might be required to produce a minimum of 4 spikes within a time window of 0.5 ms. In operational terms, this scheme could be evaluated by combining the output of the 10 fibers into a single post stimulus time histogram (PSTH) with 0.5-ms-wide bins. If any one of these bins contains more than 4 spikes, the system makes a ''stimulus present'' decision. In the forward-masking paradigm, the stimulus present decision is made if the criterion is exceeded during the time that the probe was presented. It is assumed that the model knows how to ignore events before and after the probe.

Once the number of AN fibers is specified, it is necessary to set the decision criterion value. This process is illustrated in Fig. 9(A), which shows the PSTH of 10 model HSR fibers during a single presentation of a 2AFC forward-masking paradigm stimulus. In this case the maskers were switched off and the only stimulation is a 25-dB SPL probe near the end of the second interval. Most of the activity in the PSTH is therefore spontaneous, and it can be seen that the criterion of >3 spikes/bin is not exceeded during this example of spontaneous activity. Assuming that the ideal criterion is one that is as low as possible while not subject to frequent false alarms, we may set the criterion to ''>3 spikes'' before a stimulus present decision is made.

This is only a crude representation of how a CN unit might respond. However, this simple arrangement is enough to illustrate the general principle involved without the distraction of further neural modeling. The number of fibers and the PSTH binwidth (0.5 ms) used in the evaluation are speculative but reflect a physiological scale. The criterion of >3 spikes is consistent with numbers suggested by Ferragamo *et al.* (1998) in their *in vitro* study of the response of T-stellate projection neurons in CN. Their data indicate that the number of convergent inputs may be very small and they show examples of units driven to respond with as few as four or five simultaneous AP inputs.

B. Model evaluation

In the evaluation to be described, the stimulus parameters were chosen to simulate a study of psychophysical forward masking by Jesteadt *et al.* (1982) using a 2AFC paradigm (Leavitt, 1971) with a decision rule that estimated the signal level required for 70.7% correct performance. The evaluation aims to simulate a subset of their results [see Fig. 9(C)]. The masker and probe stimuli were presented at a frequency of 4 kHz. The masker was 300 ms and the probe was 20 ms in duration. The duration of both masker and probe included 10-ms onset and offset cosine-squared ramps. When evaluating the model, the stimulus consisted of two successive 500-ms intervals; the first contained the masker with no probe and the second contained the masker followed by the probe.

The model parameters are unchanged from the previous evaluation. An example of the response of the model to the 2AFC stimuli with maskers at 80 dB SPL is given in Fig. 9(B). The model's task was to choose the interval that contained the probe. In each interval, a positive detection was made if any PSTH bin contained an above-criterion number of spikes during the probe-presentation window. A separate decision was made for each interval as to whether a probe was detected. If only one interval gave a detection decision,

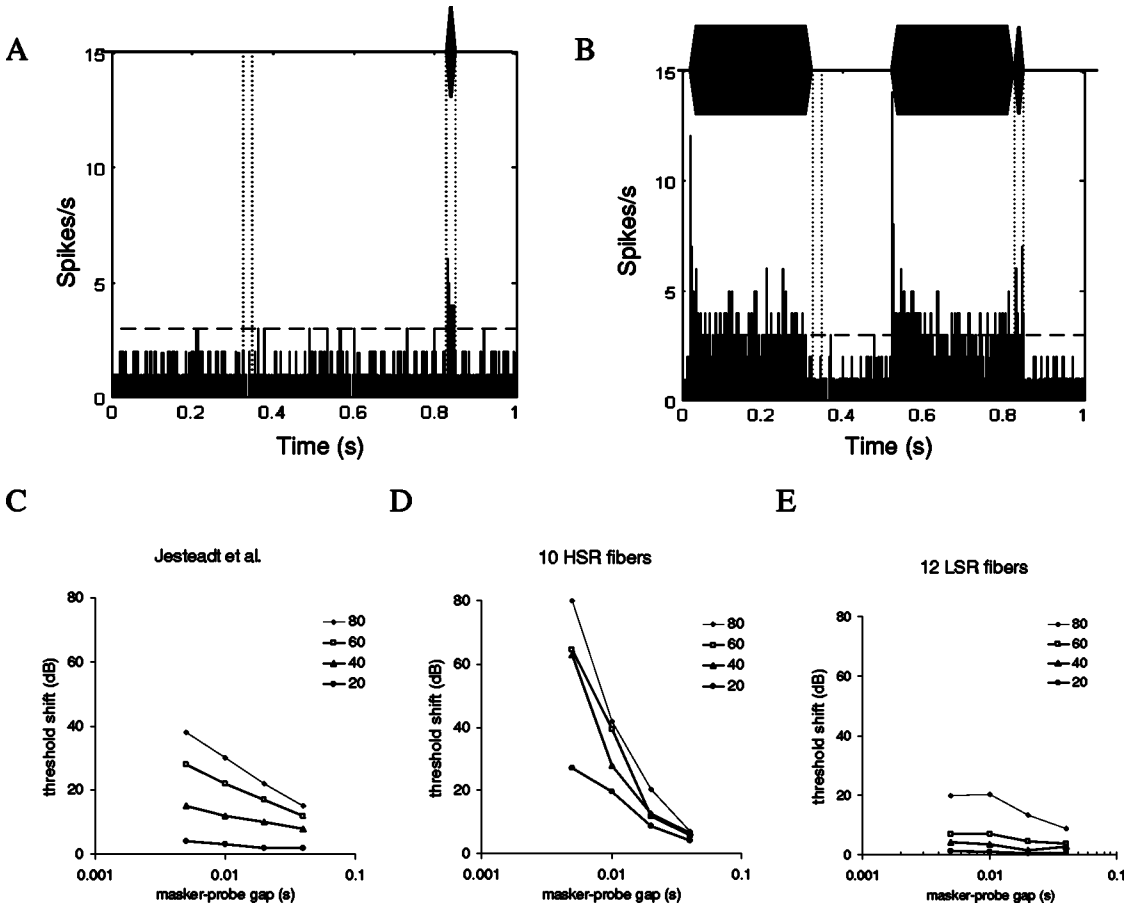


FIG. 9. Forward masking. (A) Setting the criterion for the detection of a stimulus. The PSTH is the response of 10 HSR fibers to a single presentation of the two intervals. The maskers are switched off to observe spontaneous activity. The raised activity towards the end of the PSTH is the response to a 20-dB-SPL probe tone. Vertical dotted lines indicate the no-probe and probe locations for the two measurement intervals. The criterion (dashed horizontal line) is set at a level that is rarely exceeded during spontaneous activity. (B) An example of a detection decision based on the PSTH. The maskers and probe are both 80 dB SPL. The criterion is exceeded during the second (probe) interval and the probe is detected. No detection event occurs during the first (no probe) interval. The model decision is that the probe occurred in the second interval. (C) Psychophysical forward masking data redrawn from Jesteadt *et al.* (1982) showing the threshold shift relative to the probe threshold in the absence of a masker at four different masker levels. (D), (E) Model forward masking using four masker levels and four masker–probe gaps. Threshold shifts are the difference between ''threshold after a masker'' and ''threshold without a masker.'' (D) Using 12 HSR fibers and a criterion of >3 spikes per 0.5 ms bin. (E) Using 12 LSR fibers and a criterion of >1 spike per 0.5 ms bin.

that interval was chosen. If neither or both intervals signaled a detection, a random choice was made. Unmasked thresholds were estimated on the basis of 16 trials. All thresholds are quoted as shifts relative to the unmasked threshold.

In the first version, the model consists exclusively of 10 identical HSR fibers evaluated in stochastic mode. The results are shown in Fig. 9(D). Thresholds rise substantially at short masker probe gaps, especially when high level maskers are used. Up to 70 dB of masking can be seen in the results. This masking is considerably greater than the threshold increase obtained when simulating the experiment of Relkin and Turner (<10 dB). It is larger than the maximum masking of Jesteadt *et al.* but consistent with that found elsewhere in the psychophysical studies (for example, Widin and Viemeister, 1979). Figure 9(E) shows the masking to be expected from the model if an LSR fiber is used. Because of the low spontaneous rate of firing, the criterion was reduced to >1. Much less masking is present for this fiber type at short masker–probe gaps.

While the new coincidence-detection method for measuring thresholds yields large threshold shifts, the threshold functions for both the HSR and LSR fibers are qualitatively different from those of Jesteadt *et al.* The masking recovery slopes are too steep for the HSR fiber and too shallow for the LSR fiber. However, further parameter manipulation in search of a perfect match between model and psychophysical data may not be appropriate. There are a range of reasons why the model and the human data may not match exactly. The number of fibers and combinations of fiber types may be different. In reality it is likely that the decision involves many CN units. The rate of recovery of available transmitter may be different in humans from chinchillas. The amount of BM compression may also be different between humans and the guinea pig on which the model function is based. The slopes of the functions shown in Fig. 9 are strongly influenced by the amount of compression occurring in the cochlea. The greater the compression, the wider the spacing of the thresholds at the shortest gaps and the steeper the recovery slopes. The modeler could take advantage of any of these numerous possibilities to produce a better fit, but it is doubtful whether this would serve a useful purpose and will not be pursued here. The main point to be made is that the simultaneity criterion has introduced more masking than is observed in systems based on AN spike counts.

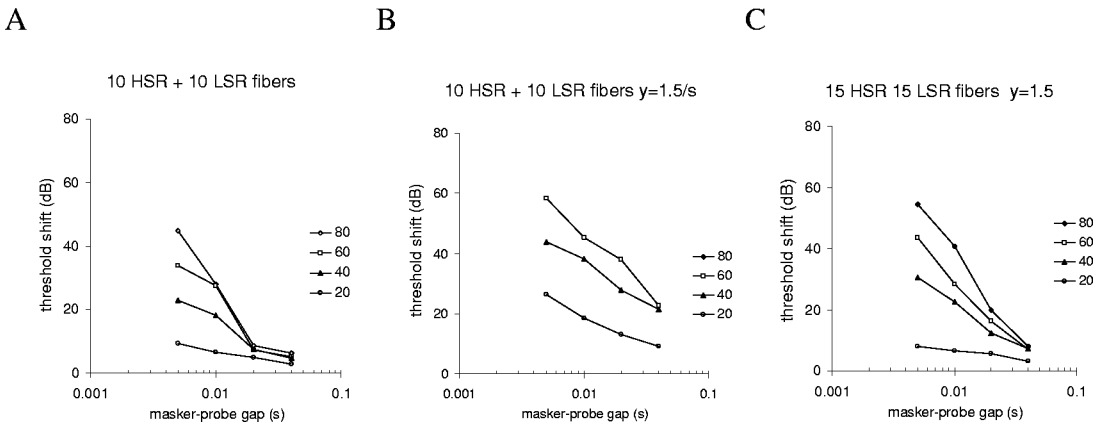


FIG. 10. Parameter manipulations that affect the shape of the forward-masking functions (mean of 10 trials). (A) Forward masking with a combination of 10 HSR and 10 LSR fibers and using a criterion of >3 spikes per bin. (B) As (A) with a slower replenishment rate ($y=1.5$). (C) As (B) with 15 HSR and 15 LSR fibers.

Nevertheless, it is relevant to note how parameter changes influence the masking functions. Figure 10(A) shows the response of a mixture of 10 HSR and 10 LSR fibers. Adding LSR fibers reduces the slope of the functions because the LSR fibers contribute to lower thresholds at shorter masker–probe gaps. However, the functions still converge at the longest gap and this is not a feature of the psychophysical data. If the replenishment rate is slowed from $y=3$ to $y=1.5$, thresholds rise, the slopes are more shallow, and recovery is clearly not complete after 40 ms [Fig. 10(B)]. When the number of fibers is increased to 15, threshold shifts are reduced for all observations.

VI. DISCUSSION

The AN model of Sumner *et al.* (2002, 2003a, and 2003b) has been shown to be able to simulate the forward-masking functions of auditory-nerve fibers as measured by Harris and Dallos (1979). The simulations became both quantitatively and qualitatively accurate after only minor changes to the value of the reprocessing and replenishment parameters. The parameter changes do not adversely disturb previously established properties of the model response. The time course of short-term adaptation is made slower but remains consistent with published animal data. Future work using the models described in Sumner *et al.* will benefit from the use of lower rates of reprocessing and replenishment.

The success of the modeling work is consistent with the idea that two separate processes (fast and slow—''reprocessing'' and ''replenishment'') are involved in the recovery of the AN response. Reprocessing is associated in the model with rapid reuptake and reprocessing of transmitter. This is active at all masker intensities. A second, slower recovery process is associated with the replenishment of transmitter lost from the synaptic region. The nature of this replenishment is unknown. However, Griesinger *et al.* (2002) have identified reuptake at remote apical sites in inner hair cells followed by transfer of transmitter material back to the basolateral synaptic region. This longer route may be a candidate for the slower replenishment process in the model.

The model was also successful in simulating the forward-masking data of Relkin and Turner (1988) in that only modest threshold shifts were found for HSR fibers when a rate measure was used in a psychophysical threshold measurement paradigm. The model showed larger shifts with LSR fibers and this was also consistent with the data. The model results support the conclusion of Relkin and Turner that a criterion for signal detection, based on single fiber AN spike counts, is unlikely to be a useful basis for explaining psychophysical measurements of forward masking made with human listeners.

A new method for target-detection was implemented and evaluated. This was based on the concept of near-simultaneous action potentials in a collection of AN fibers converging on a single unit in the CN. This approach was shown to be successful in reproducing the much larger threshold shifts observed in psychophysical forward-masking paradigms. There are two important differences between this approach and that adopted by Relkin and Turner. First, the new method uses a coincidence criterion, while Relkin and Turner used a total spike count. Second, the new method makes a separate decision for each interval, while Relkin and Turner's method employs a single decision based on the difference in the counts for the no-probe and probe intervals. This latter distinction is important because the coincidence-detection approach permits ''single trial'' (yes/no) decisions when only one stimulus is presented. The spike count method can only be used when a comparison between two stimuli is possible. For this reason, the coincidence-detection approach promises to be more generally useful in future modeling applications that reach beyond 2AFC paradigms.

The model gave useful results despite the crude implementation of the coincidence-detection decision-making mechanism using the criterion of a minimum number of spikes in a single PSTH bin. The decision mechanism was deliberately kept simple to demonstrate the basic principle. A more realistic implementation might use a model neuron whose dendritic time constant would replace the fixed-width PSTH bin. In practice, threshold decisions are likely to involve a number of coincidence detectors and the detectors will be subject to excitatory and inhibitory influences from other active units in the auditory brain stem. For these reasons, this study has stopped short of a full investigation using

a CN neural model. This is a possible line of investigation for a future project.

HSR model fibers showed more rapid recovery than those based on LSR fibers. This is consistent with the data of Relkin and Doucet (1991), who found that LSR fibers take longer to recover from adaptation than HSR. This nonintuitive observation occurs in the model despite the fact that the recovery and replenishment time parameters (x and y) are the same irrespective of fiber type. This can be explained with reference to Eq. (1). Replenishment takes place at a rate specified by the function $y[M\text{-}q_t]$; it is a function based *both* on y and on $M\text{-}q_t$. This describes a situation where the presynaptic store has only M places that a transmitter vesicle can occupy. If all but one are occupied (i.e., $M\text{-}q_t=1$), the rate of replenishment will be y. However, if many places are vacant (i.e., $M\text{-}q_t>1$), the rate of replenishment will be greater. The available transmitter pool is normally almost full for LSR fibers because the rate of release is low. On the other hand, the transmitter pool for HSR fibers typically contains many vacant places because of the high rate of release. The effective rate of recovery is therefore greater for HSR fibers than LSR fibers. The only parametric difference between HSR and LSR fibers in the model involves the rate of calcium influx at the synapse. The many differences among fiber types in functions such as rate level, thresholds, adaptation, and recovery from adaptation are all emergent properties arising from this one difference.

The success of the model in demonstrating forward masking on the same scale as psychophysiological observations should not be taken to imply that the response of a single cell at the level of the auditory nerve and the cochlear nucleus is a sufficient explanation of their results. Such a radical conclusion is not justified. We already know that efferent inhibitory mechanisms influence the cochlear nucleus response to forward masking stimuli (Shore, 1998) and the rich inhibitory networks of the brainstem and cortex are also likely to influence the full picture. Similarly, confusion and temporal uncertainty (e.g., Moore and Glasberg, 1982; Neff, 1985) can influence psychophysical forward masking in a way that suggests there is more to forward masking than the depression of AN responding. Forward masking can also be observed in cochlear implant patients where IHC physiology may not be relevant (Lim *et al.*, 1989; Shannon, 1983). Nevertheless, the results of the modeling study do indicate that adaptation and recovery from adaptation in the auditory nerve are capable of playing a substantial role in this process.

This becomes clear when a coincidence-detection approach is applied as an additional stage to the decision-making process. Coincidence detection is an almost unavoidable processing stage given what we know about the anatomy and physiology of the auditory system. It also confers the important benefit of a mechanism that distinguishes between spontaneous (uncorrelated) and driven AN activity. HSR fibers have low thresholds and are likely to be more commonly involved in decisions near absolute threshold. However, their lower thresholds are normally associated with considerable amounts of spontaneous activity. The problem of distinguishing between a stimulus-related spike and random activity is critical in this context. Coincidence detection is a simple way of achieving this goal while maintaining detection thresholds close to rate threshold. It is possible that the mechanisms necessary to distinguish between spontaneous and driven AN activity are responsible for the suboptimal performance reported by Relkin and Turner when tested using psychophysical forward-masking paradigms.

VII. CONCLUSIONS

(1) A model of the auditory periphery has been shown to be capable of simulating the data of Harris and Dallos concerning recovery from adaptation in AN fibers.
(2) The same model was able to replicate the observation of Relkin and Turner (1988) that threshold shifts following stimulation in AN fibers were smaller than those seen in psychophysical studies of forward masking.
(3) Large threshold shifts consistent with those found in psychophysics were observed when the detection mechanism was changed to one that was insensitive to spontaneous activity and based on near-simultaneous activity in parallel AN fibers.
(4) When HSR and LSR fiber types were combined, the thresholds as a function of gap and masker level most closely approximated human psychophysical forward-masking data.

ACKNOWLEDGMENTS

The authors gratefully acknowledge helpful comments made on a preliminary version of this manuscript by Enrique Lopez-Poveda, Christopher Plack, Susan Shore, and Christian Sumner. This research was supported by the Wellcome Trust and the Biotechnology and Biological Science Research council.

[1]The use of the same term has the potential to create confusion. Nevertheless, the practice has become so widespread that the term ''forward masking'' will be used here in both physiological and psychophysical contexts. The use of the term is intended to be purely descriptive of the paradigm employed and no implication concerning mechanism is intended.
[2]Owing to a proofreading error, Table I (Sumner *et al.*, 2002) gives $s1$ (displacement sensitivity) as $5E-7$, when it should be $5E-9$.

Chimento, T. C., and Schreiner, C. E. (**1990**). ''Time course of adaptation and recovery from adaptation in the cat auditory nerve neurophonic,'' J. Acoust. Soc. Am. **88**, 857–864.
Duifhuis, H. (**1973**). ''Consequences of peripheral frequency selectivity for nonsimultaneous masking,'' J. Acoust. Soc. Am. **54**, 1471–1488.
Ferragamo, M. J., Golding, N. L., and Oertel, D. (**1998**). ''Synaptic inputs to stellate cells in the ventral cochlear nucleus,'' J. Neurophysiol. **79**, 51–63.
Golding, N. L., Robertson, D., and Oertel, D. (**1995**). ''Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision,'' J. Neurosci. **15**, 3138–3153.
Griesinger, C. B., Richards, C. D., and Ashmore, J. (**2002**). ''FM 1-43 reveals membrane recycling in adult inner hair cells of the mammalian cochlea,'' J. Neurosci. **22**, 3939–3952.
Harris, D. M., and Dallos, P. (**1979**). ''Forward masking of auditory nerve fiber responses,'' J. Neurophysiol. **42**, 1083–1107.
Heil, P., and Neubauer, H. (**2001**). ''Temporal integration of sound pressure determines thresholds of auditory nerve fibers,'' J. Neurosci. **21**, 7404–7415.
Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). ''Forward masking as a function of frequency, masker level, and signal delay,'' J. Acoust. Soc. Am. **71**, 950–962.

Leavitt, H. (**1971**). ''Transformed up and down methods in psychology,'' J. Acoust. Soc. Am. **49**, 467–477.
Lim, H. H., Tong, Y. C., and Clark, G. M. (**1989**). ''Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea,'' J. Acoust. Soc. Am. **86**, 971–980.
Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). ''A computational algorithm for computing nonlinear auditory frequency selectivity,'' J. Acoust. Soc. Am. **109**, 2852–2861.
Moore, B. C. J., and Glasberg, B. (**1982**). ''Contralateral and ipsilateral cueing in forward masking,'' J. Acoust. Soc. Am. **71**, 942–945.
Moore B. C. J., Glasberg B. R., Plack C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Neff, D. (**1985**). ''Stimulus parameters governing confusion effects in forward masking,'' J. Acoust. Soc. Am. **78**, 1966–1976.
Oxenham, A. J. (**2001**). ''Forward masking: Adaptation or integration?'' J. Acoust. Soc. Am. **109**, 732–741.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). ''Spiral vos final report, Part A: The auditory filterbank,'' Cambridge Electronic Design, Contract Report (Apu 2341).
Plomp, R. (**1964**). ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.
Relkin, E. M., and Doucet, J. R. (**1991**). ''Recovery from prior stimulation. I. Relationship to spontaneous firing rates of primary auditory neurons,'' Hear. Res. **55**, 215–222.
Relkin, E. V., and Pelli, D. G. (**1987**). ''Probe tone thresholds in the auditory nerve measured by two-interval forced choice procedures,'' J. Acoust. Soc. Am. **82**, 1679–1691.
Relkin, E. M., and Turner, C. W. (**1988**). ''A re-examination of forward masking in the auditory nerve,'' J. Acoust. Soc. Am. **84**, 584–591.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction,'' Hear. Res. **12**, 1–16.
Shore, S. E. (**1998**). ''Influence of centrifugal pathways on forward masking of ventral cochlear nucleus neurons,'' J. Acoust. Soc. Am. **104**, 378–389.
Smith, R. L. (**1977**). ''Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects,'' J. Neurophysiol. **40**, 1098–1112.
Smith, R. L. (**1979**). ''Adaptation, saturation and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Sumner, C., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). ''A revised model of the inner-hair cell and auditory-nerve complex,'' J. Acoust. Soc. Am. **111**, 2178–2188.
Sumner, C., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2003a**). ''Adaptation in a revised inner-hair cell model,'' J. Acoust. Soc. Am. **113**, 893–901.
Sumner, C., O'Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (**2003b**). ''A nonlinear filter-bank model of the guinea pig cochlear nerve: Rate responses,'' J. Acoust. Soc. Am. **113**, 3264–3274.
Turner, C. W., Relkin, E. M., and Doucet, J. (**1994**). ''Psychophysical and physiological forward masking studies: Probe duration and rise-time effects,'' J. Acoust. Soc. Am. **96**, 795–800.
Westerman, L. A. (**1985**). ''Adaptation and recovery of auditory nerve responses,'' Ph.D. thesis, Syracuse University, Syracuse, NY.
Widin, G. P. and Viemeister, N. F. (**1979**). ''Intensive and temporal effects in pute-tone forward masking,'' J. Acoust. Soc. Am. **66**, 388–395.
Yates, G. K., Robertson, D., and Johnstone, B. M. (**1985**). ''Very rapid adaptation in the guinea-pig auditory nerve,'' Hear. Res. **17**, 1–12.

Transient-evoked stimulus-frequency and distortion-product otoacoustic emissions in normal and impaired ears[a)]

Dawn Konrad-Martin[b)]
VA RR&D National Center For Rehabilitative Auditory Research, Portland VA Medical Center, Portland, Oregon 97239 and Department of Otolaryngology, Head & Neck Surgery, Oregon Health and Science University, Portland, Oregon 97239

Douglas H. Keefe
Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 12 October 2004; revised 15 March 2005; accepted 16 March 2005)

Transient-evoked stimulus-frequency otoacoustic emissions (SFOAEs), recorded using a nonlinear differential technique, and distortion-product otoacoustic emissions (DPOAEs) were measured in 17 normal-hearing and 10 hearing-impaired subjects using pairs of tone pips (pp), gated tones (gg), and for DPOAEs, continuous and gated tones (cg). Temporal envelopes of stimulus and OAE waveforms were obtained by narrow-band filtering at the stimulus or DP frequency. Mean SFOAE latencies in normal ears at 2.7 and 4.0 kHz decreased with increasing stimulus level and were larger at 4.0 kHz than latencies in impaired ears. Equivalent auditory filter bandwidths were calculated as a function of stimulus level from SFOAE latencies by assuming that cochlear transmission is minimum phase. DPOAE latencies varied less with level than SFOAE latencies. The ppDPOAEs often had two (or more) peaks separated in time with latencies consistent with model predictions for distortion and reflection components. Changes in ppDPOAE latency with level were sometimes explained by a shift in relative amplitudes of distortion and reflection components. The ppSFOAE SPL within the main spectral lobe of the pip stimulus was higher for normal ears in the higher-frequency half of the pip than the lower-frequency half, which is likely an effect of basilar membrane two-tone suppression.
[DOI: 10.1121/1.1904403]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.66.Yw [BLM] Pages: 3799–3815

I. INTRODUCTION

Evoked otoacoustic emissions (OAEs) provide a noninvasive probe of cochlear micromechanics. A combination of nonlinear distortion and coherent linear reflection mechanisms has been used to explain OAE generation in cochlear models that incorporate realistic (broad and tall) basilar membrane traveling waves and randomly spaced basilar membrane discontinuities. Stimulus-frequency (SF) and click-evoked (CE) OAEs at low and moderate stimulus levels are thought to be generated by a linear coherent reflection of the forward traveling wave from a random distribution of inhomogeneities on the basilar membrane near the tonotopic peak region (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera and Guinan, Jr., 2003). At higher levels, SFOAEs have been modeled as arising from a combination of linear reflection and nonlinear distortion (Talmadge *et al.*, 2000). Experimental data is consistent with distortion contributing to SFOAEs (Shera and Guinan, 1999; Goodman *et al.*, 2003; Schairer *et al.*, 2003) and CEOAEs (Yates and Withnell, 1999) at moderate- to high stimulus levels. Distortion-product otoacoustic emissions (DP) OAEs are thought to be initiated by mechanical distortion of the basilar membrane's response to two slightly different frequencies, f_1 and f_2 (where $f_1<f_2$), near the f_2 place. The cubic distortion product, with frequency $2f_1-f_2$ (and other distortion products tuned lower in frequency than f_2), travels bidirectionally from its place of generation. One component travels basally back to the ear canal, while the other travels apically to its tonotopic location and is subsequently reflected and reverse-transmitted back to the ear canal. The distortion and reflection components thus combine to produce the DPOAE measured in the ear canal (Kim *et al.*, 1980; Kemp and Brown, 1983; Talmadge *et al.*, 1998; Shera and Guinan, 1999).

Clinical applications of OAE measurements require a better understanding of the degree to which OAE recordings match and differ from results predicted by theoretical models. In the limit of low stimulus levels in an idealized cochlea in which there is a uniform distribution of random inhomogeneities on the basilar membrane, the SFOAE or CEOAE spectrum is predicted to be the same as the evoking stimulus spectrum, apart from fine-structure (Zweig and Shera, 1995). The effect of the linear coherent reflection is described in the model using an apical reflectance seen by the forward-traveling wave, in which the apical reflectance region is the tonotopic region of the basilar membrane that dominates the generation of the reverse-traveling wave detected in the ear canal as the OAE. Random inhomogeneities along the basilar

[a)]Portions of this paper were presented at the 2002 June Meeting of the Acoustical Society of America (Konrad-Martin *et al.*, 2002) and the 2004 Winter Meeting of the Association for Research in Otolaryngology (Konrad-Martin *et al.*, 2004).

[b)]Author to whom correspondence should be addressed: VA RR&D National Center For Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Road, Portland, Oregon 97239. Electronic mail: dawn.martin@med.va.gov

membrane would produce random deviations in the SFOAE or CEOAE magnitude spectrum, which would not be encoded in the stimulus spectrum, and these random deviations are expressed in terms of random deviations in the magnitude of the apical reflectance. The spectral content of the DPOAE is influenced by interactions between the two DPOAE sources, the distortion and reflection source. Similar to SFOAEs, the DPOAE reflection source is influenced by random deviations in the magnitude of the apical reflectance (Talmadge *et al.*, 1998; Shera and Guinan, 1999), and both DPOAE sources are influenced by multiple internal reflections (Dhar *et al.*, 2002).

Experimental evidence at low and moderate stimulus levels shows complex relationships between OAE spectra and evoking stimuli. Discrepancies between predictions and experimental findings in CEOAEs in guinea pig were explained by a model that incorporates intermodulation distortion as a component of CEOAE (Yates and Withnell, 1999). Intermodulation distortion effects were not strong components of transient SFOAE in human ears, although some effects observed in transient SFOAEs were not accounted for by a simple application of the coherent reflection emission theory (Konrad-Martin and Keefe, 2003). In particular, Konrad-Martin and Keefe found that SFOAE spectra in normal ears were dominated by frequencies corresponding to the high-frequency half of the evoking tone pip. The observation that the high-side SFOAE spectrum is higher than the low-side spectrum is consistent with a within-band suppression of the lower-frequency stimulus components by adjacent higher-frequency components. If there were a nonuniform distribution of inhomogeneities along the basilar membrane producing nonuniformities in the measured spectral envelopes of SFOAE, then it would be equally likely that the SFOAE spectra averaged across a population of ears would be dominated by frequencies in the high-frequency side of the main lobe of the stimulus spectrum as by frequencies in the low-frequency side. Thus, nonuniformity in the distribution of inhomogeneities does not resolve the observed asymmetry. Averaged SFOAE spectra obtained from groups of normal and impaired ears as a function of stimulus level have not yet been examined for evidence of two-tone suppression processes. Two-tone suppression effects, if present in most normal-hearing ears, may be significant in OAE generation (Konrad-Martin and Keefe, 2003; Lukashkin and Russell, 1998). Because most SFOAE measurement techniques are based on the use of two-tone suppression to obtain a nonlinear SFOAE residual, a better understanding of SFOAE suppression at moderate levels is needed.

Cochlear models predict that reflection emissions have a latency approximately equal to the round-trip travel time between the ear canal and the tonotopic place (Kemp, 1980). Consistent with this prediction, the latency of SFOAEs elicited by tone bursts and of frequency-specific OAE components obtained by wavelet time-frequency analyses of CEOAEs were approximately twice the latency of tone-burst-evoked auditory brainstem responses (ABR) at corresponding frequencies (Neely *et al.*, 1988; Tognola *et al.*, 1997). The basilar membrane travel time included in the reflection emission latency is thought to be dominated by the velocity decrease that accompanies the forward traveling wave near the tonotopic-place region (Neely *et al.*, 1988). Greater wave velocity decreases are associated with sharper tuning of the resonant peak, consistent with the idea that the basilar-membrane mechanical transfer function is minimum phase near its peak (e.g., Zweig, 1976; de Boer, 1997). Auditory filter bandwidths have been estimated at low stimulus levels from SFOAE group delay, based on the assumed relationship between SFOAE and basilar membrane group delays (Zweig, 1991; Zweig and Shera, 1995). The resulting SFOAE measurements of equivalent rectangular bandwidth (Q_{ERB}) agree with behavioral measurements of frequency tuning based on nonsimultaneous masking measurements in normal-hearing humans from 1 to 8 kHz (Shera *et al.*, 2002).

Few studies have examined the influence of stimulus level and hearing status on the latencies of SFOAEs. Reduced cochlear amplifier gain due to increased stimulus level or impaired cochlear function shifts the peak of the traveling wave toward the cochlear base and broadens the auditory filter bandwidth (e.g., Johnstone *et al.*, 1986; Ruggero *et al.*, 1997). A relationship between SFOAE latency and hearing status (ears with either normal hearing or sensorineural hearing loss) would therefore be expected, but has not been found (Prieve *et al.*, 1996). However, SFOAE onset latencies have been found to decrease with increasing stimulus level, based on measurements in the time domain (Kemp and Brown, 1983; Norton and Neely, 1987; Neely *et al.*, 1988) and frequency domain (Prieve *et al.*, 1996). An SFOAE latency in cycles was estimated from spectral periodicity measures in the frequency domain, and shown to decrease with increasing self-suppression level (Lineton and Lutman, 2003b).

Effects of level and hearing status on CEOAE latency have also been investigated. The OAE latency from time-frequency analysis of CEOAEs was found to decrease with increasing stimulus level (Tognola *et al.*, 1997). The spectral periodicity of CEOAEs, which is thought to be a measure of cochlear tuning, was similar in healthy young adults and neonates, and absent in adult ears with noise-induced hearing loss (Avan *et al.*, 2000). Sisto and Moleti (2002) and Lucentini *et al.* (2002) reported that the latency in ears with high-frequency hearing loss (>3 kHz) was larger than the latency in normal ears, with the largest difference at 2.2 kHz, approximately one-half octave below the lower bound of the hearing loss. Moreover, Lucentini *et al.* reported that the normal-hearing ear of subjects with a unilateral hearing loss showed a similarly increased latency compared to ears of subjects with no hearing loss. In contrast to Avan *et al.* (2000), Moleti and Sisto (2003) reported that CEOAE latencies were larger in healthy neonates than adults and neonatal tuning was sharper. This conclusion was based on the assumption that the round-trip middle-ear contribution to CEOAE latency was the same (0.5 ms) in neonatal and adult ears.

The interpretation of DPOAE latency is complicated by the presence of separate distortion and reflection components, each with its own latency. The group delay of DPOAEs was measured using a swept frequency technique and found to be small for $f_2/f_1 > 1.1$ (Kemp and Brown,

1983; Brown *et al.*, 1996). These small group delays were explained by proposing that the cochlear source has an invariant frequency scaling, i.e., the DPOAE source is "wave fixed" (Kemp and Brown, 1983). For more closely spaced primaries ($f_2/f_1<1.1$), primary frequency shifts resulted in large DPOAE group delays. Rapid phase changes with changes in the DPOAE frequency were explained as arising from a generator that is fixed to a reflection site ("place fixed") near the tonotopic place (Kemp, 1986; Brown *et al.*, 1996; Knight and Kemp, 2000; Kalluri and Shera, 2001). Shera and Guinan (1999) concluded that the phase behavior exhibited by wave- and place-fixed OAE are a result of fundamentally different mechanisms, with wave-fixed OAEs produced by distortion and place-fixed OAEs produced by a linear coherent reflectance mechanism. The distortion component of the $2f_1-f_2$ DPOAE is thought to produce a group delay that is the sum of the forward travel time of the two primary tones to their place of overlap near the f_2 place, and the reverse travel time of the generated DPOAE back to the ear canal. The reflection component of the $2f_1-f_2$ DPOAE includes an additional round-trip travel time between the f_2 place of generation and the $2f_1-f_2$ tonotopic place (Tubis *et al.*, 2000b).

Numerous studies have investigated the effect of hearing loss on DPOAE latency. In one study, DPOAE group delays decreased in subjects following a temporary noise-induced hearing loss (Engdahl and Kemp, 1996). Similarly, DPOAE group delays at 4 kHz in ears that were characterized by a 4 kHz "audiometric notch" indicative of a noise-induced hearing loss were shorter than group delays in normal ears (Londero *et al.*, 2002). The majority of studies, however, have failed to find consistent effects on DPOAE group delay of permanent hearing loss (Prijs and Schoonhoven, 1997; Ramotowski and Kimberley, 1998; Kimberley, 1999; Hoth and Weber, 2001; Namyslowski *et al.*, 2001) or temporary hearing loss (Brown *et al.*, 1993). No studies in human ears have examined the effects of hearing status on DPOAE latency measured directly in the time domain.

Transient-evoked DPOAE latencies recorded in the time domain in a limited number of subjects have been reported to decrease with increasing stimulus level (Whitehead *et al.*, 1996; Konrad-Martin and Keefe, 2003). DPOAE group delays recorded in the frequency domain decreased with increasing stimulus level (Bowman *et al.*, 1997; Wable *et al.*, 1996; Kimberley *et al.*, 1993; Bowman *et al.*, 1998; Wable *et al.*, 1997), although the data are characterized by large intersubject variabilities. The reported level-dependence of DPOAE group delay may be explained by changes in the relative contributions by multiple sources to the recorded DPOAE (Stover *et al.*, 1996; Konrad-Martin *et al.*, 2001). Stover *et al.* (1996) showed that DPOAE f_1-frequency sweep data was characterized by multiple delay components when converted into an equivalent time representation using inverse fast Fourier transform (IFFT) techniques. The latencies of IFFT waveform delay components were approximately constant, but relative amplitudes varied over a range of L_1. In a later study, a suppressor tone placed just below the f_{DP} removed all IFFT waveform peaks except for the earliest peak, suggesting that the suppressed peaks were reflection components arising near the tonotopic place (Konrad-Martin *et al.*, 2001).

Examining temporal details of the transient-evoked DPOAE waveform envelope has been useful for identifying delay contributions attributed to the two DPOAE sources (Whitehead *et al.*, 1996; Talmadge *et al.*, 1999). For example, a large reflection component was predicted to cause an envelope notch near the stimulus onset in time-domain representations of DPOAE while a large distortion component would produce a notch near the stimulus offset (Talmadge *et al.*, 1999). Such notches have been observed, which shift from onset to offset with changes in L_1 in some ears, consistent with a shift in dominant source from reflection source to distortion source (see Fig. 8 in Konrad-Martin and Keefe, 2003). Except for the occurrence of such notches, it has not been possible to directly separate distortion and reflection sources in measurements of DPOAE waveform transients, and no results have been reported for short-duration stimuli such as tone pips. Using a sufficiently brief stimulus may allow the DPOAE generated by a distortion source to be separated in time from the DPOAE generated by a reflection source, and from any subsequent multiple internal reflections that may occur.

The present report extends previous work (Konrad-Martin and Keefe, 2003), which evaluated the time-frequency responses of transient-evoked SFOAE and DPOAE in selected normal-hearing subjects. The present report describes SFOAE and DPOAE responses recorded in the time domain and analyzed using a narrow-band filter at the OAE frequency of interest. This is in contrast to the time-frequency analysis based upon the cone kernel that was used in the previous work, which provided a detailed analysis of the OAE on the time-frequency plane but with unknown amounts of error in the temporal envelopes of components at the OAE frequency. The use of a more conventional narrow-band filter in the present report focuses on the temporal envelopes at the frequencies of principal OAE components. In contrast to group delay measurements in the frequency domain that must use measurements over a range of frequencies and separation techniques in order to measure the delays of multiple components, the narrow-band analysis of transient OAEs has the potential to directly measure the delays (or latencies) of multiple components in a single OAE measurement.

The present report describes data obtained in groups of normal-hearing and hearing-impaired subjects using pairs of tone pips (pp) and gated tones (gg) to evoke transient SFOAEs and DPOAEs, and continuous and gated tones (cg) as an additional stimulus to evoke transient DPOAEs. Transient stimuli with and without a steady state portion were included because it was not clear *a priori* which stimulus type would best reveal multiple OAE components. The normal-ear responses were the same as those analyzed in the previous work, whereas the impaired-ear responses have not been previously analyzed. The aims of this study were to (1) examine effects of stimulus level on time-domain measurements of SFOAE and DPOAE latency; (2) determine whether a correlation exists between hearing sensitivity and an equivalent measure of cochlear tuning (eQ_{erb}) based on

SFOAE latency; (3) determine whether ppSFOAE and ppDPOAE waveform envelopes can be used to help separate dominant DPOAE sources as a function of level; (4) test whether two-tone suppression acts differentially at higher and lower frequencies within the bandwidth of a short-duration stimulus used to evoke SFOAEs; and (5) if such a two-tone suppression effect exists in ppSFOAEs, determine whether the effect varies with level or hearing status. In addition, the methods of time-frequency analysis and narrow-band filtering are compared in selected ears.

II. METHODS

A. Subjects

Otoacoustic emissions were measured in 17 normal-hearing subjects and 10 subjects with impaired hearing. Results in each subject are reported for one ear, chosen randomly. Normal hearing was defined as pure-tone thresholds of 15 dB HL at each half-octave frequency from 0.25 to 8.0 kHz. However, the normal-hearing subjects had thresholds of 10 dB HL or better at the frequencies nearest the OAE test frequencies (3 kHz and 4 kHz). All 10 hearing-impaired subjects had pure-tone thresholds greater than 20 dB HL at 4 kHz; nine hearing-impaired subjects had pure-tone thresholds greater than 20 dB HL at 3 kHz. Fourteen female and three male subjects comprised the normal-hearing group; six female and four male subjects comprised the hearing-impaired group. Mean age in years was 17.2 (SD 8.3) for normal-hearing subjects and 56.8 years (SD 20.8) for hearing-impaired subjects (SD denotes standard deviation). All subjects had normal 226 Hz tympanometry at the time of testing. Stimulus conditions and time-frequency analyses for individual normal-ear responses were described previously (Konrad-Martin and Keefe, 2003).

B. OAE recording technique

The stimulus generation and response recording system included a signal processing board (CAC model DSP32) housed in a personal computer and controlled by custom-designed software. The probe microphone system (Etymotic Research ER-10C) included two loudspeakers to deliver stimuli and a microphone to record the pressure response. A nonlinear OAE residual waveform p_d was calculated using a single total response buffer comprised of three waveform responses, p_1, p_2, and p_{12}, sampled at 24 kHz and stored in separate 1024-sample buffers (42.7 ms in duration) by (Keefe, 1998)

$$p_d = p_1 + p_2 - p_{12}, \tag{1}$$

in which p_1 was the pressure response to a stimulus s_1 presented through loudspeaker 1, p_2 the response to a stimulus s_2 presented through loudspeaker 2, and p_{12} the response to the simultaneous presentation of s_1 and s_2 through loudspeakers 1 and 2, respectively. An average total response buffer was stored as the time average of 32 total response buffers obtained with no artifact, as determined using an artifact rejection test described in Keefe and Ling (1998).

C. Stimuli

Transient SFOAEs were obtained using s_1 and s_2 tones of equal frequency, and transient DPOAEs were recorded using different stimulus frequencies. The DPOAE stimuli were a (short-duration) pip f_2 paired with pip f_1, a gated f_2 paired with a gated f_1 (gg) or a continuous f_2 paired with a gated f_1 (cg). DPOAEs were obtained for primary frequencies $f_1 = 3.34$ kHz and $f_2 = 4$ kHz (so that $f_2/f_1 = 1.20$), which were presented at equal levels $L_1 = L_2$ (ppDPOAE) or unequal levels $L_1 = L_2 + 10$ dB (ggDPOAE and cgDPOAE). For SFOAEs, pairs of tone pips (pp) and gated tones (gg) were presented at equal levels ($L_1 = L_2$) at center frequencies of 2.7 kHz (equivalent to the f_2 used for DPOAE measurements) and 4 kHz (equal to the f_{DP} in DPOAE measurements), so that the level of the simultaneous condition was approximately $L_{12} = L_1 + 6$ dB.

To produce the tone pips, the sinusoidal signal was gated with a complete Hamming window of duration equal to 10.5 times the period of the tone. For the gated tone with nominal steady-state duration of 512 samples (21.3 ms), one-half of a Hamming window was used as an onset and offset window, with each half-window duration equal to the period of the tone being gated. The stimulus levels reported were the peak level in dB SPL of the time-domain filtered SFOAE or DPOAE (Konrad-Martin and Keefe, 2003). The OAE primary levels based on the level of an equivalent continuous tone varied in 5-dB steps from 23 to 63 dB SPL for gg and cg measurements, and from 35 to 75 dB SPL for pp measurements. Synchronous spontaneous otoacoustic emissions (SSOAE) also were measured in order to assess their contribution to SFOAE and DPOAE responses as described by Konrad-Martin and Keefe (2003).

D. Off-line analyses

Temporal envelopes of stimulus and OAE waveforms were obtained by narrow-band filtering at the SFOAE or DPOAE frequency. Each tone-pip (pp) response was filtered using a bandpass filter, which had a Kaiser finite-impulse response (FIR) design centered at the center frequency of the corresponding pp stimulus or DP frequency. Each of the pair of DPOAE primaries and DPOAE response were extracted by each of three FIR filters that differed only in center frequency. The spectral bandwidth of the filter was selected to be close to that of the pip stimulus spectrum (e.g., a pass band of 2 Hz and lower and upper reject band frequencies separated by 656 Hz were used for all pip response filters); the filter for each pip response had a group delay (2.42 ms) that was comparable to the shortest OAE component intervals that were measured. The filter envelope was within 6 dB of its peak value over a time interval of 2.88 ms. The relatively short filter duration assisted in separating individual OAE components in the time domain. Spectral ripple and the broad transition bandwidth specified above were permitted in the spectral frequency-domain passband to achieve these short durations. In particular, the filter duration was designed to be sufficiently short so that it would be possible to separate directly in the time domain the distortion and reflection sources of a ppDPOAE. The gated (gg) and continuous-

gated (cg) responses were filtered using a digital bandpass FIR filter centered at the OAE frequency (pass band of 20 Hz, reject band of 50 Hz). The group delay of the filter was 2.33 ms. A Hilbert transform was used to extract the envelope of the filtered OAE waveform, which was further processed and plotted as envelope level (in dB SPL, in which the envelope magnitude is reduced by $1/\sqrt{2}$ to convert to a root-mean-squared pressure amplitude). The same narrow-band filter was applied to coupler data to assess the distortion-plus-noise level of the measurement system. Time-domain filtered coupler responses to transient stimuli (not shown) contained low-level peaks with levels rarely exceeding -10 dB SPL. Thus, narrow-band-filtered OAE responses in ears with peaks greater than -10 dB SPL were judged to be minimally affected by artifact.

Each narrow-band filtered OAE waveform of 42.7 ms duration was classified as having an OAE present or absent based on a 6-dB ratio of time-domain signal and noise energies, and OAE latencies were calculated only for those OAE classified as present. For each pp response, the waveform buffer was separated into two halves, one containing the stimuli and OAE (plus noise) and the other containing noise alone, except for the possible (and rare) contamination from synchronous-evoked spontaneous otoacoustic emissions (SSOAE). A pp response was classified as present if the OAE energy was at least 6 dB larger in the half-buffer containing the OAE stimulus. For each gg response in which the duration of the gated response (21.3 ms) was approximately one-half the buffer duration, the waveform buffer was separated into two shorter buffers, the first containing the middle 12 ms indicative of the steady-state portion of the stimulus response (and the OAE response), and the second containing 12 ms in which noise alone was present (in addition to any SSOAE response). A gg response was classified as present if the OAE energy was at least 6 dB larger in the first buffer than the second. A cg response was classified as present using a similar 6-dB criterion to that of the gg response.

Each of the ppSFOAE and ppDPOAE latencies was calculated as the time interval between the peak envelope levels of the pip and corresponding OAE. The ggSFOAE and ggDPOAE latency was the gated tone to emission onset-time difference taken at the time in which the envelope level was 6 dB down from the steady-state level. The cgDPOAE latency was calculated using the gated f_1 stimulus in a similar manner to the ggDPOAE latency. Each measured OAE latency was classified as valid in the Results section if it exceeded a minimum criterion latency $T_{\min}$, equal to one-half the duration of the pip stimulus. $T_{\min}$ was 1.94 ms at 2.7 kHz, and 1.31 ms at 4.0 kHz. This lower bound controlled for time-domain artifact during the time that the pip stimulus was on; the possible contributions of nonlinearity in middle-ear and cochlear response in the basal region of the basilar membrane are discussed later. The same criterion was used for the gg and cg conditions to maintain consistency with the pp condition, even though the stimuli differed slightly in onset and decay times.

The ppSFOAE responses were also analyzed in the frequency domain for the presence of an asymmetry in the SFOAE spectrum, which was previously explained as an effect of basilar membrane two-tone suppression acting within the bandwidth of the pip stimulus (Konrad-Martin and Keefe, 2003). This proposed mechanism was further examined by separate measurements in the populations of normal-hearing and hearing-impaired ears. To quantify this asymmetry, the SFOAE energy, which is defined as the squared magnitude of the SFOAE pressure spectrum, was separately summed over the high-frequency and low-frequency halves of the bandwidth of the main lobe of the pip stimulus spectrum, and each energy was converted to a SPL.

TABLE I. Estimated means and confidence intervals for eQ_{ERB} and k from Shera *et al.* (2002).

	Frequency	Mean	95% confidence interval
eQ_{ERB}	2.7 kHz	17.1	16.1–18.2
	4.0 kHz	19.2	17.7–20.9
k	2.7 kHz	2.15	1.88–2.44
	4.0 kHz	2.09	1.79–2.43

Based on the assumption that the cochlear filter responsible for OAE generation is minimum phase, Shera *et al.* (2002) defined a noninvasive estimator of cochlear tuning called the equivalent rectangular bandwidth [denoted herein as $eQ_{\mathrm{ERB}}(f)$, with the leading "e" indicating an emission-based variable]. The $eQ_{\mathrm{ERB}}(f)$ is based on the SFOAE latency T in response to a tone at frequency f measured at a fixed, moderate level near 40 dB SPL. Measured in units of the number of periods of the tone, the round-trip latency is Tf, so that the forward latency to tonotopic place is $Tf/2$, assuming that the round-trip SFOAE latency is the sum of equal forward- and reverse-transmission latencies. The $eQ_{\mathrm{ERB}}(f)$ is proportional to the forward SFOAE latency with a proportionality constant $k(f)$, which is considered to vary slowly with frequency,

$$eQ_{\mathrm{ERB}}=\frac{kTf}{2}. \tag{2}$$

Shera *et al.* (2002) showed that the SFOAE-based estimator $eQ_{\mathrm{ERB}}(f)$ predicted the low-level behavioral equivalent rectangular bandwidth (denoted as Q_{ERB}) based on a forward masking paradigm. This is evidence that the noninvasive SFOAE measure of cochlear tuning accurately estimates cochlear tuning as determined behaviorally. The time-domain recorded SFOAE latencies in the present report were converted into estimates of $eQ_{\mathrm{ERB}}(f,L)$ using Eq. (2) based on the values of $k(f)$ reported by Shera *et al.* (2002). The means and confidence intervals for $eQ_{\mathrm{ERB}}(f)$ and $k(f)$ reported by Shera *et al.* are shown in Table I.

III. RESULTS

SFOAE and DPOAE responses were separately analyzed for groups of normal and impaired ears. SSOAE responses were measured to assess their contribution to the SFOAE and DPOAE responses.

Figure 1 shows an example of a pip-evoked (pp) SFOAE and a gated tone-evoked (gg) SFOAE in an im-

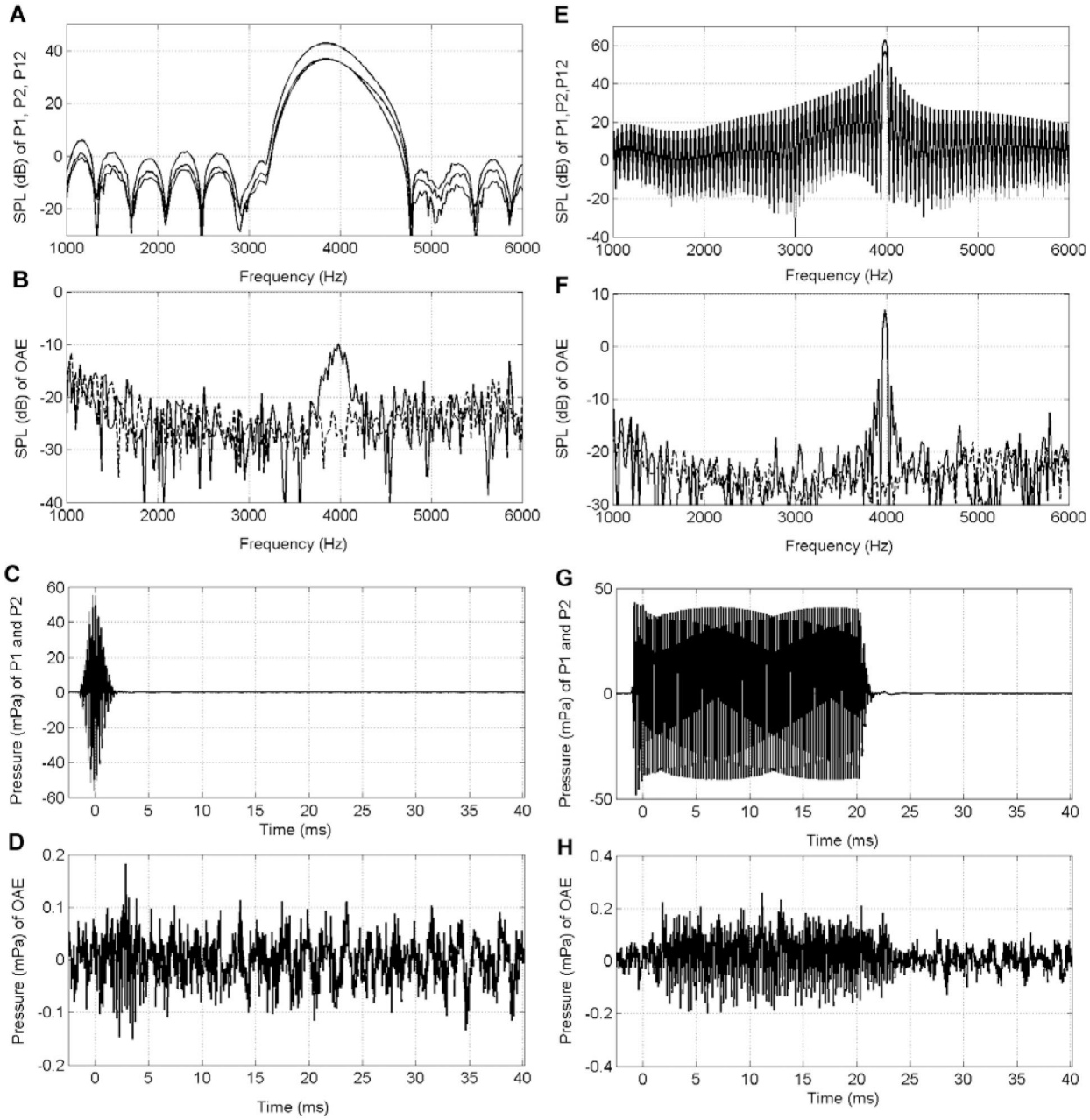


FIG. 1. Example of a pip-evoked (pp) SFOAE (left column) and a gated-tone-evoked (gg) SFOAE (right column) in an impaired ear ($f_1, f_2 = 4.0$ kHz). The equivalent continuous SPL was 66.2 dB for the pip and 63 dB for the gated tone. (A), (E) stimulus pressure spectra for p_1, p_2, and p_{12}; (B), (F) SFOAE spectrum; (C), (G) stimulus waveforms for p_1, p_2, and p_{12}; (D), (H) SFOAE waveform.

paired ear characterized by a mild flat hearing loss with a 4 kHz pure-tone threshold of 40 dB HL. The pp and gg SFOAE spectra for the stimuli centered at 4.0 kHz [Figs. 1(B) and 1(F), respectively] were similar to the corresponding stimulus spectra [Figs. 1(A) and 1(B), respectively], except that the ppSFOAE spectrum [Fig. 1(B)] was more narrow than the pip stimulus spectrum. In the time domain, the pp and gg stimulus waveforms are shown in Figs. 1(C) and 1(G), respectively, and the unfiltered OAE response waveforms in Figs. 1(D) and 1(H), respectively. In the case of ppSFOAEs, the increase in waveform amplitude indicative of the SFOAE in Fig. 1(D) occurred at a latency exceeding $T_{\min}$ and was large relative to the noise floor of the narrow-band-filtered response (not shown). Thus, this impaired ear had measurable SFOAE responses.

A. Effect of mild to moderate hearing loss on valid SFOAE latency recordings

Transient-evoked SFOAE responses were compared for normal-hearing subjects and subjects with mild to severe cochlear hearing loss. Figure 2 shows SFOAE latency at 2.7 kHz (left column) and 4.0 kHz (right column) plotted as a function of the pure-tone threshold at the closest audiometric test frequencies (3 kHz and 4 kHz, respectively). Latencies shown were elicited by tone pips (top row) and gated tones (bottom row) similar to responses depicted in Figs. 1(A)–

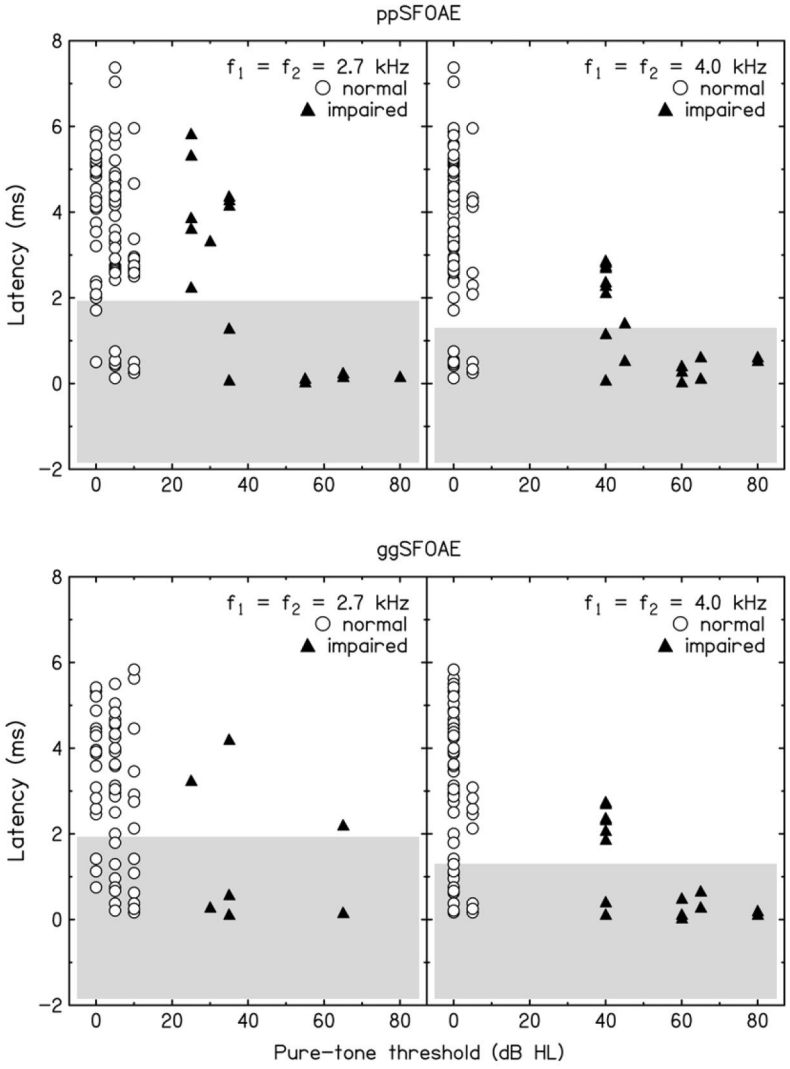


FIG. 2. Effect of hearing status on SFOAE latency at two frequencies. SFOAE at 2.7 kHz (left column) and 4.0 kHz (right column) are plotted as a function of pure-tone threshold at the closest test frequency (3 kHz and 4 kHz, respectively). Data are collapsed across stimulus level. Rows correspond to the particular class of stimulus used (top row=*pp*SFOAE data; bottom row=*gg*SFOAE data), which are depicted in Fig. 1. Circles represent data from normal ears, and triangles represent data from impaired ears. The gray shading identifies ears with latencies shorter than $T_{\min}$.

1(D) and 1(E)–1(H), respectively. For each transient stimulus class and frequency, data are collapsed across the range of primary levels. Circles represent data from normal ears; triangles represent data from impaired ears. In Fig. 2 and subsequent figures, the measured latencies shorter than $T_{\min}$ are shown with gray shading and the remaining (''valid'') OAE latencies are shown with white background.

Responses with signal to noise ratio (SNR) exceeding the 6 dB criterion were obtained for a wide range of audiometric threshold levels. In particular, SFOAEs were present in Fig. 2 for thresholds up to and including 80 dB HL. However, valid latencies were obtained only in subjects with pure-tone thresholds below about 45 dB HL. Because latencies shorter than $T_{\min}$ may be related to artifact (see section D of the Discussion), the latencies shorter than $T_{\min}$ were not considered reliable indicators of cochlear function and were excluded from statistical analyses.

Subjects in the data set with impaired hearing and valid latencies had moderate hearing loss (40 to 45 dB HL) at 4.0 kHz, and mild hearing loss (25 to 35 dB HL) at 2.7 kHz, suggesting that any differences between groups with normal and impaired-hearing would be greater at 4.0 kHz compared to 2.7 kHz. Consistent with the fact that the *pp* stimuli had a higher peak SPL than the *gg* stimuli, valid *pp*SFOAE latencies were observed more often than valid *gg*SFOAE latencies in both groups of impaired and normal ears. The proportions of normal and impaired subjects with valid *pp* and *gg* SFOAE latencies and SNR≥6 dB are given in Table II. SFOAEs were present more often in normal-hearing than impaired-hearing subjects.

B. Effect of stimulus level and hearing status on SFOAE latency

Figure 3 shows the effect of stimulus level (L_1) on *pp*SFOAE (top row) and *gg*SFOAE (bottom row) latencies for groups of normal (open circles) and impaired ears (filled triangles). The left-hand axis represents the measured range of latencies. Responses to 2.7 kHz and 4.0 kHz stimuli are

TABLE II. Proportion of subjects with valid tone-pip (*pp*) evoked SFOAE and gated-tone (*gg*) evoked SFOAE. SNR>6 dB and latencies greater than $T_{\min}$=1.94 ms at 2.7 kHz and 1.31 ms at 4.0 kHz.

	*pp*SFOAE		*gg*SFOAE	
	2.7 kHz	4.0 kHz	2.7 kHz	4.0 kHz
Normal	16/17	16/17	14/17	14/17
Impaired	5/9	4/10	3/9	2/10

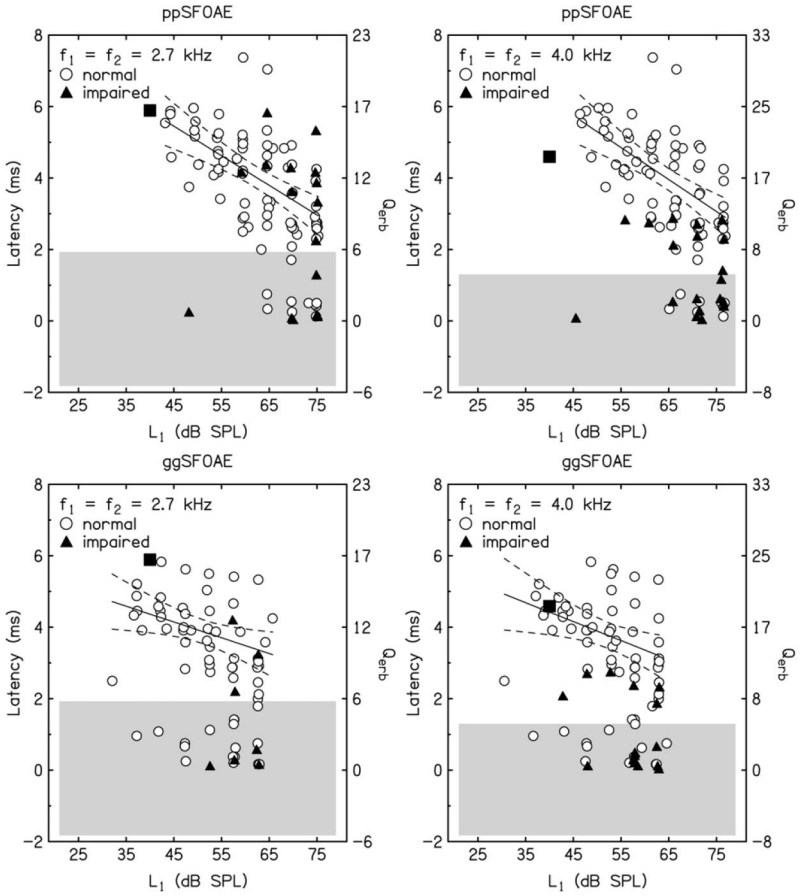


FIG. 3. Effect of stimulus level on SFOAE delay (left-hand axis) and eQ_{ERB} (right-hand axis). Latency data from Fig. 2 are replotted as a function of stimulus level. Linear fits and corresponding 95th confidence intervals are shown for normal data that meet latency criteria ($>T_{\mathrm{min}}$). The mean eQ_{ERB} reported by Shera *et al.* (2002) are indicated by the filled squares at 40 dB SPL. Format is the same as for Fig. 2.

presented in left and right columns, respectively. Linear fits (solid lines) and associated 95% confidence intervals (dashed lines) are also shown for data obtained in normal ears and meeting minimum-latency and SNR criteria. The effect of level on eQ_{ERB} is also shown in Fig. 3, in which the right-hand axis represents the range of eQ_{ERB}.

Separate linear regression analyses were performed for valid 2.7 kHz and 4.0 kHz ppSFOAE and ggSFOAE data in normal ears in order to assess the effect of level on measured latencies. In normal ears at both 2.7 and 4.0 kHz, increasing L_1 significantly reduced ppSFOAE latencies ($p<0.001$) and ggSFOAE latencies ($p<0.01$). A necessary consequence of this result with that of Eq. (2) is that increasing L_1 reduced SFOAE estimates of cochlear tuning (eQ_{ERB}). The mean SFOAE latency and eQ_{ERB} reported by Shera *et al.* (2002) using continuous-tone stimuli are indicated in Fig. 3 by the filled squares at 40 dB SPL. Over a 30-dB range from 45 to 75 dB SPL, the regression estimates of ppSFOAE latency and eQ_{ERB} decreased at higher levels by 2.5 ms (46%) at 2.7 kHz and by 2.7 ms (46%) at 4.0 kHz. The rates of change of latency with level appear to be greater for ppSFOAE responses (top row) than ggSFOAE responses (bottom row). This apparent difference between SFOAE types, however, was related to a single low-latency (2.5 ms) outlier in ggSFOAEs produced in the same ear for 2.7 and 4.0 kHz gated tones presented at $L_1 \approx 30$ dB SPL. Removal of the outliers resulted in ggSFOAE latency changes (40% at 2.7 kHz and 49% at 4.0 kHz with a 30 dB change in L_1) that were comparable to ppSFOAE latency changes.

At 4 kHz, the SFOAE latencies in normal ears tended to be larger than the corresponding latencies in impaired ears. Separate unpaired (independent samples) *t*-tests with unequal variances assumed were used to test the hypothesis that the differences in the mean latencies obtained in normal and impaired ears at 4.0 kHz occurred by chance. The results of these analyses indicate that the group of subjects with hearing impairment had significantly shorter mean ppSFOAE ($p<0.001$) and ggSFOAE latencies ($p<0.001$) compared to normal-hearing subjects. However, the mean differences were likely due to the fact that it was rarely possible to measure latencies in impaired ears at lower stimulus levels due to an insufficient SNR. It was possible to measure valid ppSFOAE latencies at more than two L_1 in several impaired ears. One ear at 2.7 kHz had valid data at four L_1 levels and two ears at 4.0 kHz had valid data at three and five L_1 levels, respectively. Level-dependent changes in latency were absent in these impaired-ear data at 2.7 kHz, and were absent in the second of these two ears at 4.0 kHz. Analysis of individual waveform envelopes (not shown) showed that the high-latency ppSFOAE outliers in Fig. 3 (data points near 7 ms) were due to multiple internal reflections in an ear with a strong SSOAE at 2.7 kHz. Compared to the earliest peak in the ppSFOAE time-domain representation, the second peak was larger for those stimulus levels that elicited the long latencies in Fig. 3. In such cases, the latency taken at the ppSFOAE SPL maximum corresponds to the delay of an OAE after one complete internal reflection within the cochlea.

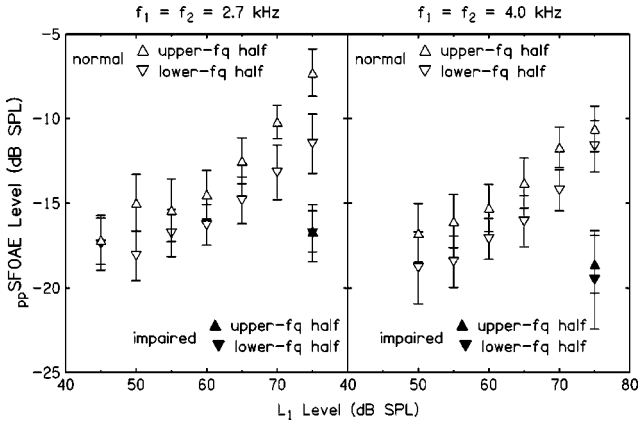


FIG. 4. Effect of stimulus level on SFOAE asymmetry at two frequencies. The SPL in the lower- (down-pointing triangles) and upper-frequency halves (up-pointing triangles) of the pip stimulus spectrum is plotted at 2.7 kHz (left) and 4.0 kHz (right) as a function of L_1 for normal (open symbols) and impaired ears (closed symbols).

C. Effect of stimulus level on SFOAE spectral asymmetry

Konrad-Martin and Keefe (2003) reported that *pp*SFOAE bandwidths were more narrow than the eliciting stimulus spectrum in a number of normal ears, and the *pp*SFOAE response components corresponded primarily to the upper-frequency portion of the stimulus passband. In contrast, the *pp*SFOAE in the impaired ear in Fig. 1(B) shows a more narrow bandwidth than the stimulus spectrum in Fig. 1(A), but there is no pronounced asymmetry in the *pp*SFOAE spectrum around its center frequency (4 kHz), as was observed in normal ears.

Results are shown in Fig. 4 for the group means and standard errors (SE) of the *pp*SFOAE SPL for normal (open symbols) and impaired ears (filled symbols) at 2.7 kHz (left panel) and 4.0 kHz (right panel) in the lower-(triangles pointing downward) and upper-frequency (triangles pointing upward) halves of the main lobe of the pip stimulus spectrum. These means and SEs are plotted as a function of pip stimulus level L_1 for all conditions in which three or more responses were present. The mean SPL of the upper-frequency half of the *pp*SFOAE spectrum was consistently larger than the mean SPL of the lower-frequency half in normal ears. The effect was most pronounced for 2.7 kHz pips at the higher stimulus levels. The *pp*SFOAE SPLs were lower in ears with mild to moderate hearing loss compared to normal ears at $L_1 = 75$ dB SPL. For both the 2.7 and 4.0 kHz pips, this largest L_1 was the only level at which three or more impaired ears had a *pp*SFOAE present. Additionally, the *pp*SFOAE spectral asymmetry present in normal ears was absent in the impaired-ear data [consistent with Fig. 1(B)].

D. Effect of stimulus level and hearing status on DPOAE latency

Using filtered DPOAE waveforms elicited by various transient stimulus types, DPOAE latencies at 2.7 kHz were measured directly in the time domain as a function of L_1 and hearing status. Figure 5 shows the latencies measured in normal (circles) and impaired ears (triangles) for narrow-band filtered *pp*DPOAE [Fig. 5(A)], *gg*DPOAE [Fig. 5(B)], and *cg*DPOAE [Fig. 5(C)] waveforms. Linear regression analyses on normal ears with valid responses showed that the effect of level on latency was significant at the 0.05 alpha level for *cg*DPOAE ($p = 0.004$) waveforms, but not for *pp*DPOAE or *gg*DPOAE waveforms.

The DPOAE latencies were typically less than 4 ms and tended to vary less with level than did the SFOAE latencies (see Fig. 3, left column) measured at the same frequency. Similar to phase-slope group delays, the components from DPOAE reflection and distortion sources are unresolved in the results presented in Fig. 5. Details of the narrow-band filtered *pp*DPOAE waveform envelope were further examined in order to determine whether the distortion and reflection sources could be separately measured. Pip-evoked DPOAE responses with SNR$\geq$6 dB were obtained in 13 of 17 normal-hearing ears and 2 of 10 impaired ears. However, waveform peaks (≥ -6 dB SPL) in these ears often had latencies less than $T_{\min}$.

Figure 6 shows the effect of L_1 on waveform envelopes of narrow-band filtered *pp*DPOAE obtained in three normal-hearing subjects. Data for each subject are presented in separate columns and L_1 decreases with descending row number in the figure. Shown on the horizontal axis is the time delay

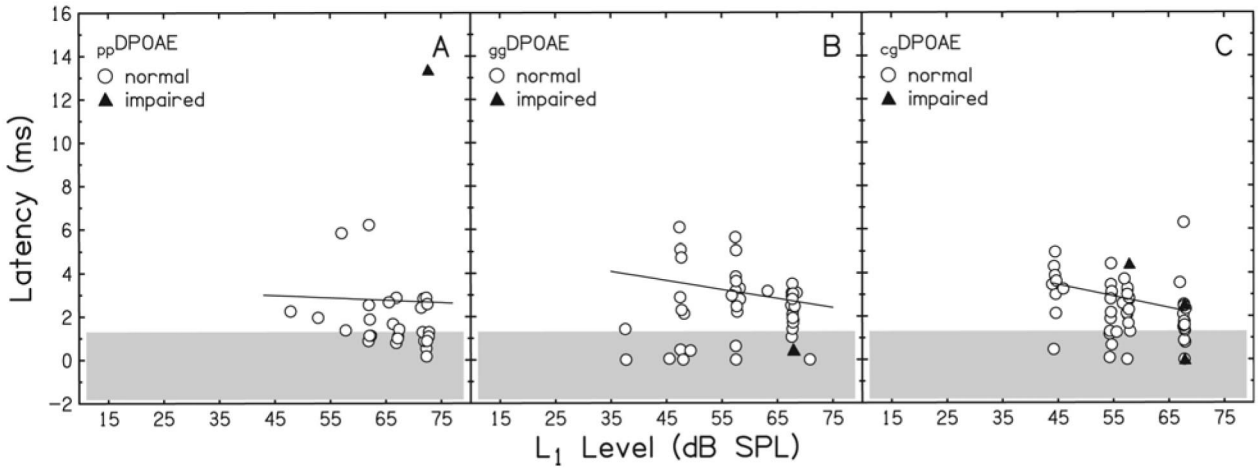


FIG. 5. Effect of stimulus level on DPOAE latency. (A) *pp*DPOAE. (B) *gg*DPOAE. (C) *cg*DPOAE. Format is the same as for Figs. 2 and 3.

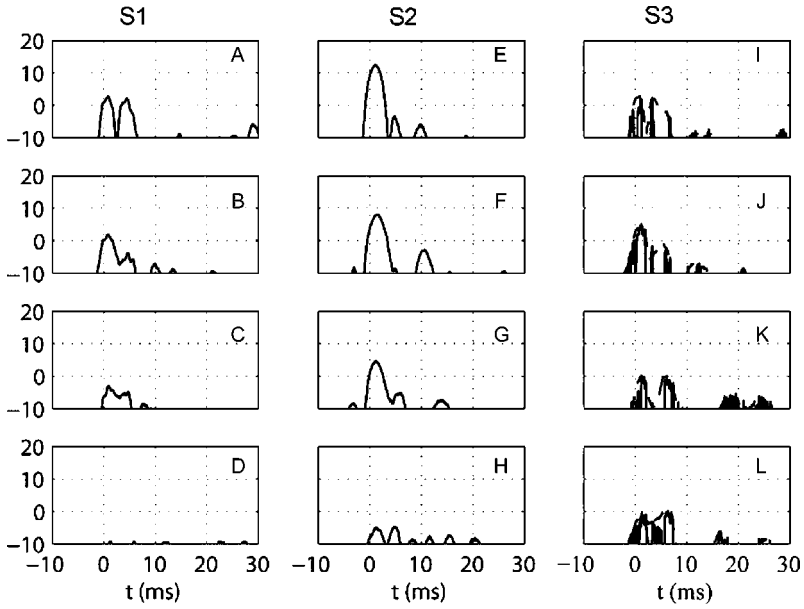


FIG. 6. The envelope SPLs of the *pp*DPOAE waveforms after narrow-band filtering around the center frequency of the DP frequency (2.7 kHz) are shown for normal-hearing subjects S1 [left column panels (A)–(D)], S2 [panels (E)–(H)], and S3 [dashed lines in panels (I)–(L)] as functions of time. The envelope SPLs of the *pp*DPOAE waveforms after TFR processing are shown for subject S3 [solid lines in panels (I)–(L)]. The stimulus level in each column decreases with increasing row number from top to bottom.

of the OAE waveform relative to the time of the maximum SPL of the narrow-band-filtered stimulus. Slightly less than half of the temporal *pp*DPOAE envelopes recorded in normal ears (7/15) contained a single narrow peak in time; the remaining recordings (8/15) had two or more peaks.

For subject S1, Figs. 6(A)–6(D) show a set of DPOAE envelopes that each contain a pair of prominent narrow peaks. The earliest *pp*DPOAE peak (Peak 1) dominated the waveform response at each L_1. The latency of Peak 1 was short (0.6 ms) at the highest L_1, and remained fairly constant across L_1. At the highest L_1, a second peak (Peak 2), which was similar in SPL to Peak 1, had a slightly longer latency of approximately 4.3 ms.

For subject S2, The set of waveform envelopes [Figs. 6(E)–6(H)] have in common an early peak at 1.0–1.26 ms. This earliest peak (Peak 1) decreased in level from 12.2 to −5.2 dB SPL with only minor changes in latency (0.26 ms) as L_1 decreased over a 16 dB range from 73 to 57 dB SPL. At the highest L_1, a low-level second peak was present with a 4.6 ms latency. The second peak was absent in Fig. 6(F), but reappeared with greater relative level at the lower stimulus levels shown in Figs. 6(G) and 6(H). The third peak near 10 ms in Figs. 6(E) and 6(F) may be a first-order reflection of the distortion component, however, its latency relative to Peak 1 is long [8.74 ms in Fig. 6(E), 9.18 ms in Fig. 6(F)].

For subject S3, the dashed-line set of *pp*DPOAE responses [Figs. 6(I)–6(L)] resemble those shown for subject S1 [Figs. 6(A)–6(D)]. The latency of Peak 1 increased only about 0.25 ms as L_1 decreased from 72 down to 57 dB SPL. In contrast, the latency of Peak 2 was increased by approximately 2.2 ms (from 3.7 ms to 5.9 ms) as L_1 decreased, and its envelope level changed nonmonotonically. At some stimulus levels [e.g., $L_1 = 63$ dB SPL in Fig. 6(K)], the sources were of equal strength, whereas at other levels the distortion source was dominant [$L_1 = 67$ dB SPL in Fig. 6(J)] or the reflection source was slightly dominant [$L_1 = 57$ dB SPL in Fig. 6(L)]. The hypothesis based on modeling and previous measurements is that the reflection source dominates at lower stimulus levels, which was the case at the lowest L_1 (57 dB SPL), and the distortion source dominates at higher stimulus levels, which was the case at the second highest level (67 dB SPL). The hypothesis was not strongly supported at the highest level [73 dB SPL in Fig. 6(I)].

The *pp*DPOAE envelopes at f_{DP} were also extracted for comparison using a time-frequency representation (TFR) described in Konrad-Martin and Keefe (2003), and are shown in the solid-line curves of Figs. 6(I)–6(L). The narrow-band-filtered envelopes are more smoothly varying than the TFR envelopes because of the effect of the impulse response of the narrow-band filter, but the peaks in the envelopes nonetheless coincide in latency. There is a tradeoff between the increased time-frequency resolution of the TFR envelopes and the presence of signal-processing artifact introduced in the calculation of TFR envelopes that is absent in the corresponding calculation of narrow-band envelopes (see discussion and references in the Appendix of Konrad-Martin and Keefe, 2003). Despite these differences, the narrow-band-filtered and TFR envelopes of the *pp*DPOAE responses agreed in this and other cases in identifying the times over which the envelopes were of moderate to large level. This provides additional evidence that the two-peak structure of these temporal envelope responses is a property of the cochlear-generated response.

With the understanding that the DPOAE latencies may vary considerably depending on which of the distortion and reflection sources is dominant, the previous group data showing *pp*DPOAE latencies were reexamined to evaluate whether the overall variability may have been increased due to this effect of source dominance. Figure 7(A) shows the *pp*DPOAE latencies from normal-hearing ears, replotted from Fig. 5(A). The latencies of the components of Peak 1 (asterisks) and Peak 2 (filled circles) in the valid filtered *pp*DPOAE waveforms are plotted separately in Fig. 7(B) for comparison with Fig. 7(A). These results suggest that a contributor to the variability in the original *pp*DPOAE latency data [Fig. 7(A)] is the fact that the distortion source was dominant in some ears and the reflection source was dominant in other ears. Moreover, the level-dependence of

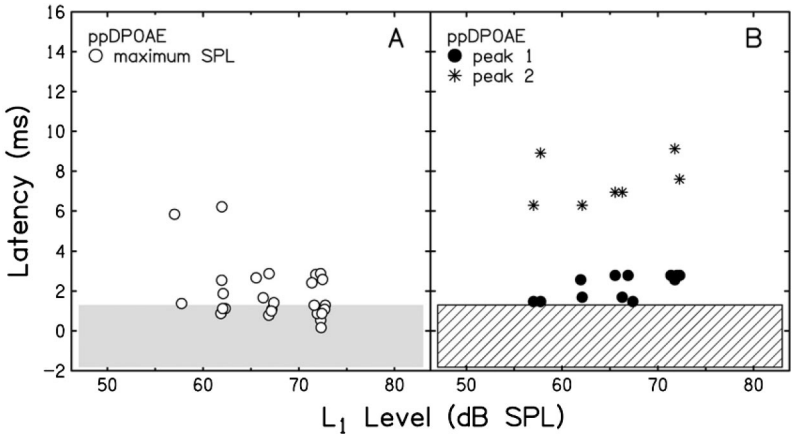


FIG. 7. (A) *pp*DPOAE latencies from normal ears replotted from Fig. 5(A). These latencies are for the maximum OAE SPL. (B) Corresponding Peak 1 and Peak 2 latencies from valid time domain waveforms.

DPOAE latency was explained in several ears by a maximum *pp*DPOAE SPL that corresponded to the earliest peak (Peak 1) at lower L_1 and to the second peak in time (Peak 2) at higher L_1, consistent with an overall shift in dominance to the distortion source at higher L_1.

Table III shows the number of ears for each DPOAE type meeting the criteria that a DPOAE was present with an interpretable latency. DPOAEs in most normal ears did not meet these criteria for L_1 below approximately 45 dB SPL (see Fig. 5 for the lowest L_1 at which the criteria were met for *pp*DPOAE, *gg*DPOAE, and *cg*DPOAE). In only one or two of the impaired ears did the DPOAE waveforms meet the criteria for any of the DPOAE types. Overall, the prevalence of DPOAEs in impaired ears was less than the prevalence of SFOAEs.

IV. DISCUSSION

A. Summary

In a previous report, transient-evoked SFOAE and DPOAE responses were measured in individual normal-hearing subjects (Konrad-Martin and Keefe, 2003). Group analyses in the present study show the effects of level on spectral-temporal properties of OAE signals measured in the same normal-hearing subjects and in an additional 10 hearing-impaired subjects. The following results were obtained. (1) Latencies decreased with increasing stimulus level for SFOAE elicited in normal ears at 2.7 and 4.0 kHz using pairs of tone pips (*pp*) and gated tones (*gg*). (2) Valid impaired-ear latencies were similar or shorter compared to normal-ear latencies at equal SPL; however, the number of impaired-ear responses was insufficient to determine whether *pp*SFOAE and *gg*SFOAE latencies differed significantly at any one L_1 in impaired compared to normal ears. The observed SFOAE latency variation with level and hearing status suggests that Q_{ERB} varies under the same conditions. Thus, the possibility that transient-evoked SFOAE may provide a rapid, noninvasive measure of cochlear tuning warrants further study. (3) Latencies of envelope peaks of *pp*DPOAE and *pp*SFOAE provided evidence for two $2f_1-f_2$ DPOAE sources (a distortion source near the f_2 place and a reflection source at the DP place). Increasing the stimulus level produced little or no changes in the latencies attributed to a distortion source. Rather, variation in *pp*DPOAE latencies corresponded to variation in the dominant generator source. (4) In normal, but not impaired ears, the *pp*SFOAE spectrum was dominated by energy above the center frequency, particularly at the highest stimulus levels tested. Basilar membrane two-tone suppression appears to be involved in OAE responses to such stimuli as tone pips and may be important for interpreting OAE elicited using more realistic sounds common to speech and music.

TABLE III. Proportion of subjects with valid tone-pip (*pp*) evoked DPOAE and gated-tone (*gg*) evoked DPOAE and continuous-f_2, gated-f_1. SNR≥6 dB and latencies greater than T_{min}=1.31 ms.

	*pp*DPOAE	*gg*DPOAE	*cg*DPOAE
Normal	9/17	15/17	16/17
Impaired	1/10	0/10	2/10

B. Frequency-specific time-domain recorded SFOAE latency

The present research reports measurements of SFOAE latency *T* over a range of stimulus levels *L*. The valid time-domain recorded SFOAE latencies obtained for moderate-level tone pips and gated tones presented at 2.7 and 4.0 kHz (see Fig. 3) were consistent with SFOAE latencies at moderate levels in which a continuous suppressor tone was paired with an intermittent probe tone (Brass and Kemp, 1991) but contrast with other reports of transient-SFOAE data with longer latencies (Neely *et al.*, 1988; Prieve *et al.*, 1996). The SFOAE latencies in the present study were in the range of frequency-domain SFOAE group delays obtained using continuous tones presented at a fixed moderate (40 dB SPL) level (Shera *et al.*, 2002). There was general accord between these measured SFOAE latencies and model predictions based on the WKB theory of cochlear mechanics and the coherent linear-reflection mechanism. For example, a SFOAE analysis similar to the reflection-source component of DPOAEs in the model formulation of Tubis *et al.* (2000a, 2000b) based on parameter values from Talmadge *et al.* (1998) predicts SFOAE latencies of 6.4 ms and 4.3 ms for 2.7 kHz and 4.0 kHz, respectively. These are in the range of the mean values in Fig. 3 (see also Konrad-Martin and Keefe, 2003 for further discussion of the WKB theory). The

measured values in Fig. 3 show considerable inter-ear variability that tends to obscure the variability due to changes in probe frequency, and that also limit the ability to assess the relative precision in the parameter values of Talmadge *et al.* (1998).

The SFOAE latency-based eQ_{ERB} values for moderate-level tone pips and gated tones presented at 2.7 and 4.0 kHz (see Fig. 3) were similar to psychophysical Q_{ERB} values estimated using nonsimultaneous behavioral masking for probe levels near threshold (Shera *et al.*, 2002). This similarity is due to the fact that the SFOAE latencies measured for moderate-level pips and gated tones were similar to those measured for low-level continuous tones by Shera *et al.*, and both studies calculated the eQ_{ERB} variable in the same manner. The argument that tone-pip evoked SFOAEs may provide a noninvasive estimate of the auditory filter bandwidth stems from predictions that SFOAE latency is determined by basilar membrane traveling wave delay at low levels of stimulation (Zweig, 1991; Zweig and Shera, 1995) for a basilar membrane modeled as a set of minimum phase filters (Zweig, 1976; de Boer, 1997). The varying SFOAE latencies obtained in the present study over a range of stimulus levels suggest an effect of cochlear nonlinearity.

C. Stimulus level and hearing status affect SFOAE latency and eQ_{ERB}

The approach used here to estimate cochlear tuning assumes that the value of *k*, which Shera *et al.* (2002) calculated based on measurements in normal ears at a SFOAE probe level of 40 dB SPL, is also applicable for deriving estimates of $eQ_{\mathrm{ERB}}(f,L)$ in ears with hearing loss and over a range of probe levels. Thus, changes in round-trip latency due to probe level or hearing status are assumed to be proportional to changes in cochlear tuning. In the present study, level-dependent shifts in SFOAE latency were observed in normal ears. The present data show a reduction in SFOAE latency with a 30 dB increase in probe-tone level from 45 to 75 dB SPL (48% at 2.7 and 4.0 kHz). This reduction is slightly larger, but likely similar within measurement precision, compared to the reductions (38%) observed and predicted previously for tone-pip-evoked SFOAEs (Neely *et al.*, 1988).

Recorded OAEs in the present study were considered "valid" if their SNR exceeded 6 dB and if their latency exceeded a criterion value (T_{min}). These valid OAEs, which are shown in Figs. 2, 3, and 5 (data on white background), were judged to be of cochlear origin. For valid SFOAEs obtained at the highest L_1 levels tested (i.e., 75 dB SPL for *pp*SFOAE and 63 dB SPL for *gg*SFOAE), the majority of latencies were reduced to approximately 2–3 ms compared to longer delays at lower L_1. Similar level-dependent shifts in latency have been reported (Schairer *et al.*, 2003; Kalluri *et al.*, 2004; Schairer *et al.*, 2004). This level-dependent shift is consistent with a decrease in group delay of the apical cochlear reflectance in the coherent reflection emission theory that would result, at least qualitatively, from a level-dependent broadening of the cochlear traveling wave peak with increasing level (Zweig and Shera, 1995). The observed level-dependent latency shift is also consistent with a basal shift in the maximum region of the cochlear excitation pattern with increasing level, which would shift the region in which coherent linear reflection occurred toward the base as well as broaden the region. In addition, SFOAEs obtained at high L_1 may also be generated by a nonlinear distortion source at the primary frequency place (Brass and Kemp, 1993; Talmadge *et al.*, 2000; Goodman *et al.* 2003, 2004).

Lineton and Lutman (2003a) constructed a model of cochlear tuning based on broadband self-suppression of SFOAEs and predicted a reduction "of around 50% for a suppression of the normalized SFOAE of around 20 dB." While their measured tuning bandwidths were less in absolute terms than predicted bandwidths, their frequency-domain experiments revealed a reduction in their measure of cochlear tuning from 17 to 10 as the suppressed SFOAE level varied over a range of 18 dB due to variations in stimulus level from 14 to 50 dB SPL (Lineton and Lutman, 2003b). Thus, their measure of cochlear tuning was reduced by 41% over a 36 dB range of stimulus levels. Although the methodologies differ greatly between studies, the results of the present study (a 48% increase in eQ_{ERB} over a 30 dB range of stimulus levels) are similar to Lineton and Lutman (2003b), especially taking account of the fact that a large intersubject variability was observed in each study. One difference in the studies is that valid *pp*SFOAE latencies, and thus eQ_{ERB} values, were obtained in 16 of 17 normal-hearing ears in the present study, but SFOAE bandwidth data were reported in only 9 of 20 normal-hearing ears in Lineton and Lutman (2003b).

In the present study, SFOAEs were usually absent in impaired ears at lower stimulus levels, as would be expected. Data in ears with mild to moderate cochlear hearing loss in which SFOAE responses were present showed that mean absolute SFOAE latencies were similar to (at 2.7 kHz) or shorter (at 4.0 kHz) than SFOAE latencies obtained in normal ears. It follows that eQ_{ERB} values in impaired ears, when present, were similar to or less than eQ_{ERB} values in normal ears. In contrast to results in normal ears, there was little or no evidence for a level-dependent shift in SFOAE latencies in impaired ears.

Predictions based on the coherent reflection theory for SFOAE generation include a correspondence between SFOAE latencies and basilar membrane traveling wave delays, which are in turn related to cochlear tuning for a basilar membrane that is minimum phase. Whereas the observed SFOAE latency changes with hearing status and L_1 are in the predicted direction (i.e., shorter latencies are expected for impaired ears compared to normal ears and at increased L_1 in normal ears), previous reports of CEOAE have shown longer latencies in impaired ears (Sisto and Moleti, 2002; Lucentini *et al.*, 2002). Therefore, the absolute SFOAE latencies in impaired-ears at 2.7 kHz that fall within the range of normal values in Fig. 3 (left column) warrant further discussion. An important issue is to determine whether changes in SFOAE-based estimates of cochlear tuning $eQ_{\mathrm{ERB}}(f,L)$ due to probe level and hearing status mirror changes in cochlear tuning that have been observed using psychoacoustical measures in humans.

There is consensus in the psychoacoustical literature that

frequency selectivity is reduced in individuals with impaired hearing (see Moore, 1998 for review). In a recent study comparing frequency tuning in normal and hearing-impaired ears with thresholds from 20 to 50 dB HL, Baker and Rosen (2002) estimated auditory filter shapes from psychophysical growth of masking data, which were obtained using simultaneously presented notched-noise maskers and probe tones presented at 2 kHz over a range of levels. Q_{ERB} values have been calculated from the ERB data in Fig. 5 of Baker and Rosen (2002), which also included normal-ear ERB data of Rosen *et al.* (1998). Absolute Q_{ERB} in ears with hearing loss was poorer compared to normal ears at low to moderate levels, but differences were less apparent at high stimulus levels where tuning is relatively broad in normal ears. In normal ears, Q_{ERB} decreased about 36% for a 30 dB increase in probe-tone level from 45 to 75 dB SPL. Consistent with the literature (Stelmachowicz *et al.*, 1987; Murnane and Turner, 1991), ears with mild to moderate hearing impairment showed little or no variation in auditory filter bandwidth as a function of probe-tone level. Previous studies demonstrated that hearing loss is associated with changes in ERB only when thresholds increase above 30 to 40 dB HL (Laroche *et al.*, 1992; Glasberg and Moore, 1986).

The present study included OAE responses in impaired ears with a similar hearing loss characteristic to the subjects in the Baker and Rosen study. In particular, the group of impaired ears at 2.7 kHz had pure-tone thresholds at 3 kHz between 25 and 35 dB HL (see Fig. 2). Based on eQ_{ERB} as the SFOAE measure of frequency selectivity, the frequency selectivity was similar between normal and impaired ears at the higher stimulus levels at which it was possible to measure valid SFOAEs. To this extent, these SFOAE results agree with the behavioral results of Baker and Rosen (2002). The frequency selectivity diminished in normal ears with increasing level in both behavioral (Rosen *et al.*, 1998) and SFOAE measurements. In agreement with the psychoacoustical literature, the SFOAE measure of frequency selectivity in impaired ears had little or no variation at higher stimulus levels (see Fig. 3). The main limitation in the present SFOAE results is the inability to measure SFOAEs at low to moderate stimulus levels in impaired ears.

It is also possible that in impaired ears with detectable OAEs, the source of the OAE may be altered in a complex manner from the source in a normal ear by cochlear lesions basal to the tonotopic location. In studies on the effects of experimentally induced cochlear lesions, a transient increase in OAE level preceded an OAE level reduction for lesions due to aminoglycoside ototoxicity (Kakigi *et al.*, 1998) and noise over-exposure (Harding and Bohne, 2004). Such increases in OAE level might be expected to be accompanied by increases in OAE latencies following cochlear injury. However, OAE latencies were not measured in these studies.

Results in ears with mild losses, which demonstrate normal tuning or latencies in some studies, but not in others may reflect variability in specific mechanisms underlying threshold elevation. The correlation between the degree of hearing loss and frequency tuning is expected to be high for hearing loss in which outer hair cell damage and/or loss is the dominant underlying pathology. However, a poorer correlation between hearing loss and tuning is expected for an etiology that includes dysfunction of inner hair cells (or auditory nerve fibers) (Moore *et al.*, 1999). Taken together, SFOAE results in the present study on effects of level and hearing status on auditory filter shape are similar to psychoacoustical results.

The present results suggest that further research using transient-evoked SFOAE and DPOAE in impaired ears is warranted. This study used a nonlinear residual technique for SFOAE measurement, in which at least some saturation of SFOAE growth is necessary for SFOAE extraction, and complete saturation of SFOAE growth is required in order to extract the entire SFOAE. Such techniques preclude SFOAE measurements at stimulus levels near threshold in normal ears and in impaired ears where basilar membrane response growth is linear. It might be useful to adjust the relative levels of the transient stimuli eliciting a DPOAE response to be chosen in accord with the findings of Kummer *et al.* (1998).

D. Origin of short-latency SFOAE

In contrast to SFOAE latencies elicited by low to moderate stimulus levels, some SFOAE latencies elicited by moderate- to high-level transients in the present study were apparently inconsistent with the reflection source hypothesis, inasmuch as many of the latencies were too short to be considered valid. The proportion of OAE responses with latencies near 0 ms increased with increasing stimulus level. This trend can be seen in the ''invalid'' group data shown with a gray background in Fig. 3 (for SFOAE) and Fig. 5 (for DPOAE). One example is the *pp*SFOAE data at both probe frequencies that are clustered between 0 and 1 ms in Fig. 3 for moderately high level tones ($L_1 \approx 63$ to $L_1 \approx 77$ dB SPL). Similar effects also are evident in the DPOAE group data shown in Fig. 5, but are more pronounced for *gg*DPOAEs and *cg*DPOAEs than for *pp*DPOAEs.

Potential sources for these short-latency SFOAEs include (1) artifact due to nonlinearities in the stimulus, (2) a distortion component in middle-ear transmission (Konrad-Martin and Keefe, 2003; Guinan *et al.*, 2003), (3) a distortion source or sources located basal to the tonotopic place (Brass and Kemp, 1993; Siegel, 2004), and (4) a reverse-transmission path for the OAE back through the cochlear fluid (Ren, 2002). Stimulus artifact was judged as unlikely, because short-latency ''SFOAEs'' were not observed in coupler recordings based on the 6 dB SNR criterion. It is possible that acoustic reflex (AR) shifts due to middle-ear muscle effects might explain OAE latencies occurring near 0 ms, as has been suggested by Guinan *et al.* (2003) in an interpretation of recorded SFOAEs. Possibilities (2) and (3) are each consistent with the observation of short-duration SFOAEs but it is difficult to separate in time a hypothetical source at the basal end of the cochlea from a hypothetical source within the middle ear. The presence of short-latency SFOAEs at low stimulus levels, in which AR shifts would be unlikely, suggests that a distortion component in the middle-ear cannot account for all of these data. The argument that short OAE latencies are evidence for the dominance of a cochlear-fluid path from the place of generation back to the ear canal would also need to explain the long latencies ob-

served at low stimulus levels, which are consistent with current theories of round-trip propagation on the basilar membrane. While stimulus artifact and middle-ear nonlinearity appear unlikely explanations for at least some of the present data, the present findings are insufficient to identify which of the latter two cochlear mechanisms may be involved.

E. Sources of variance in SFOAE latency

The ranges of ''valid'' absolute latencies were similar for *pp*SFOAE latencies and *gg*SFOAE onset latencies (compare Fig. 3 top and bottom rows). Complicated onsets and offsets are likely sources of variability in the long-latency *gg*SFOAE measurements and are consistent with individual observations of *gg*SFOAEs with dips and peaks in the envelope. Such dips and peaks were interpreted previously as resulting from temporal interactions between a direct (zero-order) emission and higher-order reflections (Konrad-Martin and Keefe, 2003), and their presence makes it difficult to assign a latency to the zero-order *gg*SFOAE component. In contrast to the gated tones used to elicit *gg*SFOAE, the band-limited impulses used to elicit *pp*SFOAE were sufficiently brief to allow the zero-order SFOAE to be separated in time from any subsequent multiple internal reflections. The high-latency outliers in the upper panels of Fig. 3 were attributable to first-order reflections with slightly larger SPL than the zero-order OAE components. These relatively large-amplitude first-order reflections were observed only in ears with SSOAEs near the OAE test frequency(ies). Since spontaneous OAE are thought to be self-sustaining standing waves (Kemp, 1979; Bialek and Wit, 1984; Burns *et al.*, 1998; Shera, 2003) requiring multiple reflections of a cochlear signal from apical (tonotopic place) and basal (oval window) reflection sites, it is not surprising that an ear with evidence of multiple internal reflections would also produce a SSOAE near the test frequency (also see Lineton and Lutman, 2003b).

F. Spectral composition of transient SFOAE

Konrad-Martin and Keefe (2003) observed spectral fine structure that was consistent within an ear for a variety of stimulus types (e.g., for clicks, brief tone pips, and longer-duration gated tones), but the spectral fine structure varied across ears. Based on analyses in a few ears, Konrad-Martin and Keefe (2003) also observed an asymmetry in the *pp*SFOAE spectrum with larger SFOAE energy in the high-frequency half of the pip spectrum than in the low-frequency half. They hypothesized that this effect was due to the same two-tone suppression mechanism described by Brass and Kemp (1993), namely, that high-frequency tones are more effective at suppressing SFOAEs than low-frequency tones. In the present study, this asymmetry was further examined by separate measurements in the populations of normal-hearing and hearing-impaired ears.

For pip-evoked SFOAEs, the bandwidth of the evoked *pp*SFOAE tended to be smaller than the bandwidth of the stimulus (as in Fig. 1). An asymmetry in level was observed in the mean *pp*SFOAE spectra in a group of normal ears at 2.7 kHz and 4.0 kHz (Fig. 4), and is consistent with a within-band suppression of the lower-frequency OAE components by the slightly higher-frequency components. The magnitude of the asymmetry increased with increasing stimulus level at 2.7 kHz and was absent from the mean impaired-ear data. These SFOAE asymmetries interpreted using two-tone suppression are consistent with previous findings that suppression increases with increases in the masker level when measured psychophysically (Gifford and Bacon, 2000; Moore and Vickers, 1997; Oxenham and Plack, 1998) and at the level of single auditory nerve fibers (Delgutte, 1996).

G. Frequency-specific time-domain measurements of DPOAE latency

At the highest levels tested, the observed DPOAE latencies with mean values near 2–3 ms (see Fig. 5) were similar to published DPOAE onset latencies. For example, the DPOAE onset latencies at $f_2=4$ kHz including probe travel time were approximately 2.8 ms for gated tones presented at $L_1=L_2=75$ dB SPL (Martin *et al.*, 1998), and the mean DPOAE phase-slope group delay was 3.4 ms for tones presented at $L_2=60$ and $L_1=65$ dB SPL (Moulin and Kemp, 1996).

The *pp*DPOAE waveforms often produced two or more narrow peaks in time (see, for example, the three normal-ear responses in Fig. 6). Individual *pp*DPOAE peaks were narrow compared to *pp*SFOAE peaks and limited by the temporal resolution of the narrow-band filter. The initial *pp*DPOAE peak, Peak 1, ranged from 0.6 ms to 2.8 ms. After ears with latencies less than $T_{\min}$ were removed to form the group of ears with valid DPOAE responses, the mean Peak 1 latency was 2.2 ms (SD 0.6 ms) and the mean Peak 2 latency was 7.4 ms (SD 1.2 ms). The *pp*DPOAE Peak 1 and Peak 2 latencies [Fig. 7(B)] agreed with model predictions for distortion and reflection components (3.0 ms and 7.1 ms, respectively), based on long-wavelength cochlear-model calculations similar to those described in Tubis *et al.* (2000a) and tested in Konrad-Martin and Keefe (2003). These predicted latencies are within the measurement precision and the lack of certainty regarding the physiological parameters in the cochlear model. These *pp*DPOAE data are consistent with the hypothesis that there are two $2f_1-f_2$ DPOAE sources (one near the f_2 place and one at the $2f_1-f_2$ place).

The *cg*DPOAE latencies also varied with stimulus level [Fig. 5(C)]. However, the analysis of the *pp*DPOAE waveforms (such as those shown in Fig. 6) suggests that the latency of the distortion component changed only slightly (<0.3 ms) with level and these changes were small relative to within-group variability for a particular level (Fig. 7). For high *pp*DPOAE latency values, Peak 2 (reflection component) more often corresponded to the maximum DPOAE SPL than did Peak 1 [compare Fig. 7(B) to Fig. 7(A)]. These data suggest that DPOAE latency changes with level in the *cg*DPOAE responses were mainly due to changes in the relative contributions of multiple sources (Stover *et al.*, 1996; Konrad-Martin *et al.*, 2001). One advantage of *pp*DPOAE responses over *gg*DPLOAE, *cg*DPOAE and continuous DPOAE responses is that it is possible to directly separate the distortion and reflection components in the time domain.

H. Comparison of time-domain DPOAEs analyzed using narrow-band filtering and TFRs

There was general accord between the SFOAE and DPOAE latencies measured at low levels that are presented here for narrow-band filtered responses and those that were measured using TFR analyses for transient OAEs (Konrad-Martin and Keefe, 2003). While only a single example was plotted [Fig. 6(C)], a similar pattern was observed in all transient SFOAE and DPOAE envelopes that were compared. The good correspondence between measurements validates the TFR technique for exploring OAE responses elicited by stimuli more complex than a pure tone.

V. CONCLUSIONS

Transient-evoked SFOAE and DPOAE measurements were useful for testing model predictions of OAE latency. Latencies of transient-evoked SFOAE decreased with increasing stimulus level and were insensitive to hearing status at high stimulus levels. The data at low and moderate levels mainly support the theory of coherent reflection as the dominant mechanism of SFOAE generation in human ears, but two-tone suppression effects were also present. However, the presence of short-latency SFOAEs in some ears support the existence of an additional cochlear source of SFOAE generation. The time-domain method of measuring pip-evoked SFOAE latencies over a range of stimulus levels may provide a fast and noninvasive indirect measure of cochlear tuning. Narrow-band filtering of pip-evoked DPOAE waveforms allowed temporal separation of DPOAE distortion and reflection sources.

ACKNOWLEDGMENTS

This research was supported by the VA Rehabilitation R&D Service (E3239V) and by the NIH (R01 DC003784 and T32 DC00013). The authors thank Jeffrey L. Simmons for assistance in data collection, and Robert H. Withnell for valuable comments on the paper.

Avan, P., Wit, H. P., Guitton, M., Mom, T., and Bonfils, P. (**2000**). ''On the spectral periodicity of transient-evoked otoacoustic emissions from normal and damaged cochleas,'' J. Acoust. Soc. Am. **108**, 1117–1127.

Baker, R. J., and Rosen, S. (**2002**). ''Auditory filter nonlinearity in mild/moderate hearing impairment,'' J. Acoust. Soc. Am. **111**, 1330–1339.

Bialek, W., and Wit, H. P. (**1984**). ''Quantum limits to oscillator stability: Theory and experiments on acoustic emissions from the human ear,'' Phys. Lett. **104A**, 173–178.

Bowman, D. M., Brown, D. K., Eggermont, J. J., and Kimberley, B. P. (**1997**). ''The effect of sound intensity on $f1$-sweep and $f2$-sweep distortion product otoacoustic emissions phase delay estimates in human adults,'' J. Acoust. Soc. Am. **101**, 1550–1559.

Bowman, D. M., Eggermont, J. J., Brown, D. K., and Kimberley, B. P. (**1998**). ''Estimating cochlear filter response properties from distortion product otoacoustic emission (DPOAE) phase delay measurements in normal hearing human adults,'' Hear. Res. **119**, 14–26.

Brass, D., and Kemp, D. T. (**1991**). ''Time-domain observation of otoacoustic emissions during constant tone stimulation,'' J. Acoust. Soc. Am. **90**, 2415–2427.

Brass, D., and Kemp, D. T. (**1993**). ''Suppression of stimulus frequency otoacoustic emissions,'' J. Acoust. Soc. Am. **93**, 920–939.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). ''Two sources of acoustic distortion products from the human cochlea,'' J. Acoust. Soc. Am. **100**, 3260–3267.

Brown, A. M., Williams, D. M., and Gaskill, S. A. (**1993**). ''The effect of aspirin on cochlear mechanical tuning,'' J. Acoust. Soc. Am. **93**, 3298–3307.

Burns, E. M., Keefe, D. H., and Ling, R. (**1998**). ''Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **103**, 462–474.

De Boer, E. (**1997**). ''Cochlear models and minimum phase,'' J. Acoust. Soc. Am. **102**, 3810–3813.

Delgutte, B. (**1996**). ''Physiological models for basic auditory percepts,'' in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York).

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). ''Multiple internal reflections in the cochlea and their effect on DPOAE fine structure,'' J. Acoust. Soc. Am. **112**, 2882–2897.

Engdahl, B., and Kemp, D. T. (**1996**). ''The effect of noise exposure on the details of distortion product otoacoustic emissions in humans,'' J. Acoust. Soc. Am. **99**, 1573–1587.

Gifford, R. H., and Bacon, S. P. (**2000**). ''Contributions of suppression and excitation to simultaneous masking: effects of signal frequency and masker-signal frequency relation,'' J. Acoust. Soc. Am. **107**, 2188–2200.

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.

Goodman, S. S., Withnell, R. H., De Boer, E., Lilly, D. J., and Nuttall, A. L. (**2004**). ''Cochlear delays measured with amplitude-modulated tone-burst-evoked OAEs,'' Hear. Res. **188**, 57–69.

Goodman, S. S., Withnell, R. H., and Shera, C. A. (**2003**). ''The origin of SFOAE microstructure in the guinea pig,'' Hear. Res. **183**, 7–17.

Guinan, Jr., J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). ''Medial olivocochlear efferent reflec in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs,'' J. Assoc. Res. Otolaryngol. **4**, 521–540.

Harding, G. W., and Bohne, B. A. (**2004**). ''Comparison of noise-induced DPOAE temporary level shift with detailed histopathology,'' Association for Research in Otolaryngology, Abstract No. 174.

Hoth, S., and Weber, F. N. (**2001**). ''The latency of evoked otoacoustic emissions: its relation to hearing loss and auditory evoked potentials,'' Scand. Audiol. **30**, 173–83.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). ''Basilar membrane measurements and the traveling wave,'' Hear. Res. **22**, 147–153.

Kakigi, A., Hirakawa, H., Harel, N., Mount, R. J., and Harrison, R. V. (**1998**). ''Comparison of distortion-product and transient evoked otoacoustic emissions with ABR threshold shift in chinchillas with ototoxic damage,'' Auris Nasus Larynx **35**, 223–232.

Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: a test of the two-mechanism model for DPOAE generation,'' J. Acoust. Soc. Am. **109**, 622–637.

Kalluri, R., and Shera, C. A. (**2004**). ''The relationship between Teoaes and SFOAEs at low stimulus levels,'' Association for Research in Otolaryngology, Abstract No. 873.

Keefe, D. H. (**1998**). ''Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence,'' J. Acoust. Soc. Am. **103**, 3489–3498.

Keefe, D. H., and Ling, R. (**1998**). ''Double-evoked otoacoustic emissions: II. Intermittent noise rejection, calibration and ear canal measurements,'' J. Acoust. Soc. Am. **103**, 3499–3508.

Kemp, D. T. (**1979**). ''The evoked cochlear mechanical response and the auditory microstructure; Evidence for a new element in cochlear mechanics,'' in *Models of the Auditory System and Related Signal Processing Techniques*, edited by M. Hoke and E. de Boer (Scandanavian Audiology Supplements, Stockholm), Vol. 9, pp. 35–47.

Kemp, D. T. (**1980**). ''Towards a model for the origin of cochlear echoes,'' Hear. Res. **2**, 533–548.

Kemp, D. T. (**1986**). ''Otoacoustic emissions, travelling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.

Kemp, D. T., and Brown, A. M. (**1983**). ''An integrated view of cochlear mechanical nonlinearities observable from the ear canal,'' in *Cochlear Mechanics*, edited by E. de Boer and M. A. Viergever (Delft U.P., Delft), pp. 75–82.

Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). ''Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure,'' J. Acoust. Soc. Am. **67**, 1704–1721.

Kimberley, B. P. (**1999**). ''Applications of distortion-product emissions to an otological practice,'' Laryngoscope **109**, 1908–1918.
Kimberely, B. P., Brown, D. K., and Eggermont, J. J. (**1993**). ''Measuring human cochlear traveling wave delay using distortion product emission phase responses,'' J. Acoust. Soc. Am. **94**, 1343–1350.
Knight, R. D., and Kemp, D. T. (**2000**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed $f1$, $f2$ area study,'' J. Acoust. Soc. Am. **107**, 457–473.
Konrad-Martin, D., and Keefe, D. H. (**2003**). ''Time-frequency analyses of transient-evoked stimulus-frequency and distortion-product otoacoustic emissions: Testing cochlear model predictions,'' J. Acoust. Soc. Am. **114**, 2021–2043.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). ''Sources of DPOAEs revealed by suppression experiments and IFFTs in normal ears,'' J. Acoust. Soc. Am. **109**, 2862–2879.
Konrad-Martin, D., Simmons, J., and Keefe, D. H. (**2002**). ''Spectral-temporal analysis of OAEs in normal-hearing humans: Latencies, fine structure, and multiple internal reflections,'' J. Acoust. Soc. Am. **111**, 2356(A).
Konrad-Martin, D., Simmons, J. L., and Keefe, D. H. (**2004**). ''Transient SFOAE and DPOAE in normal and impaired ears: Effects of stimulus level on latency and self-suppression,'' Association for Research in Otolaryngology, Abstract No. 27, p. 108.
Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the $2f1-f2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3344.
Laroche, C., Hetu, R., Tran Quoc, H., Josserand, B., and Glasberg, B. (**1992**). ''Frequency selectivity in workers with noise-induced hearing loss,'' Hear. Res. **64**, 61–72.
Lineton, B., and Lutman, M. E. (**2003a**). ''Modeling the effect of suppression on the periodicity of stimulus frequency otoacoustic emissions,'' J. Acoust. Soc. Am. **114**, 859–870.
Lineton, B., and Lutman, M. E. (**2003b**). ''The effect of suppression on the periodicity of stimulus frequency otoacoustic emissions: Experimental data,'' J. Acoust. Soc. Am. **114**, 871–882.
Londero, A., Bonfils, P., and Avan, P. (**2002**). ''Magnitudes and phases of human distortion-product otoacoustic emissions at $2f(1)-f(2)$ against $f(2)/f(1)$: effects of an audiometric notch,'' Hear. Res. **167**, 46–56.
Lucentini, M., Moleti, A., and Sisto, R. (**2002**). ''On the detection of early cochlear damage by otoacoustic emission analysis,'' J. Acoust. Soc. Am. **111**, 972–978.
Lukashkin, A. N., and Russell, I. J. (**1998**). ''A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer,'' J. Acoust. Soc. Am. **103**, 973–980.
Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). ''Locus of generation for the $2f1-f2$ vs $2f2-f1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations,'' J. Acoust. Soc. Am. **103**, 1957–1971.
Moleti, A., and Sisto, R. (**2003**). ''Objective estimates of cochlear tuning by otoacoustic emission analysis,'' J. Acoust. Soc. Am. **113**, 423–429.
Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).
Moore, B. C. J., Plack, C. J., and Oxenham, A. J. (**1999**). ''Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**, 2761–2778.
Moore, B. C. J., and Vickers, D. A. (**1997**). ''The role of spread excitation and suppression in simultaneous masking,'' J. Acoust. Soc. Am. **102**, 2284–2290.
Moulin, A., and Kemp, D. T. (**1996**). ''Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General characteristics,'' J. Acoust. Soc. Am. **100**, 1617–1639.
Murnane, O., and Turner, C. W. (**1991**). ''Growth of masking in sensorineural hearing loss,'' Audiology **30**, 275–285.
Namyslowski, G., Morawski, K., Urbaniec, P., Trybalska, G., and Lisowska, G. (**2001**). ''Latencies of the $2f1-f2$ distortion product otoacoustic emissions measured using a phase-gradient method in young people, in the elderly and in people exposed to noise,'' Scand. Audiol. Suppl. **52**, 121–125.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**, 652–656.
Norton, S. J., and Neely, S. T. (**1987**). ''Tone-burst evoked otoacoustic emissions from normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1860–1872.
Oxenham, A. J., and Plack, C. J. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **104**, 3500–3510.
Prieve, B. A., Gorga, M. P., and Neely, S. T. (**1996**). ''Click-and tone-burst evoked otoacoustic emissions in normal-hearing and hearing-impaired ears,'' J. Acoust. Soc. Am. **99**, 3077–3086.
Prijs, V. F., and Schoonhoven, R. (**1997**). ''Mechanical delays in deteriorated cochleas,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 53–58.
Ramotowski, D., and Kimberley, B. (**1998**). ''Age and the human cochlear traveling wave delay,'' Ear Hear. **19**, 111–119.
Ren, T. (**2002**). ''Longitudinal pattern of basilar membrane vibration in the sensitive cochlea,'' J. Neurosci. **99**, 17101–17106.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**). ''Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adults ears,'' J. Acoust. Soc. Am. **114**, 944–966.
Schairer, K. S., Fitzpatrick, D., Goodman, S., Ellison, J. C., and Keefe, D. H. (**2004**). ''Using Level-Dependent Latencies to Identify Dominant SFOAE Sources,'' American Auditory Society meeting.
Shera, C. A. (**2003**). ''Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves,'' J. Acoust. Soc. Am. **114**, 244–262.
Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shera, C. A., and Guinan, Jr., J. J. (**2003**). ''Stimulus-frequency-emission group delay: a test of coherent reflection filtering and a window on cochlear tuning,'' J. Acoust. Soc. Am. **113**, 2762–2772.
Siegel, J. (**2004**). ''Otoacoustic emissions: Have we been barking up the wrong tree,'' Association for Research in Otolaryngology, Abstract No. 1309.
Sisto, R., and Moleti, A. (**2002**). ''On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears,'' J. Acoust. Soc. Am. **111**, 297–308.
Stelmachowicz, P. G., Lewis, D. E., Larson, L. L., and Jesteadt, W. (**1987**). ''Growth of masking as a measure of response growth in hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 1881–1887.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.
Talmadge, C. L., Long, G. R., Rubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). ''Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure,'' J. Acoust. Soc. Am. **108**, 2911–2932.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). ''Time-frequency distributions of click-evoked otoacoustic emissions,'' Hear. Res. **106**, 112–122.
Tubis, A., Talmadge, C. L., and Tong, C. (**2000a**). ''Modeling the temporal behavior of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **107**, 2112–2127.

Tubis, A., Talmadge, C. L., Tong, C., and Dhar, S. (**2000b**). ''On the relationships between the fixed-$f1$, fixed-$f2$, and fixed-ratio phase derivatives of the $2f1-f2$ distortion product otoacoustic emission,'' J. Acoust. Soc. Am. **108**, 1772–1785.

Wable, J., Collet, L., Berger-Vachon, C., and Chery-Croze, S. (**1997**). '' $2f1-f2$ distortion product otoacoustic emission latency: changes with frequency and level of primaries,'' Audiology **36**, 72–82.

Wable, J., Collet, L., and Chery-Croze, S. (**1996**). ''Phase delay measurements of distortion product otoacoustic emissions at $2f1-f2$ and $2f2-f1$ in human ears,'' J. Acoust. Soc. Am. **100**, 2228–2235.

Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). ''Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency,'' J. Acoust. Soc. Am. **100**, 1663–1679.

Yates, G. K., and Withnell, R. H. (**1999**). ''The role of intermodulation distortion in transient-evoked otoacoustic emissions,'' Hear. Res. **136**, 49–64.

Zweig, G. (**1976**). ''Basilar membrane motion,'' in *Cold Spring Harbor Symposia on Quantitative Biology, Volume XL, 1975* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 619–633.

Zweig, G. (**1991**). ''Finding the impedance of the organ of Corti,'' J. Acoust. Soc. Am. **89**, 1229–1254.

Zweig, G., and Shera, C. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination

Joshua G. W. Bernstein[a] and Andrew J. Oxenham
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

(Received 5 April 2004; revised 28 February 2005; accepted 15 March 2005)

Fundamental frequency (f_0) difference limens (DLs) were measured as a function of f_0 for sine- and random-phase harmonic complexes, bandpass filtered with 3-dB cutoff frequencies of 2.5 and 3.5 kHz (low region) or 5 and 7 kHz (high region), and presented at an average 15 dB sensation level (approximately 48 dB SPL) per component in a wideband background noise. Fundamental frequencies ranged from 50 to 300 Hz and 100 to 600 Hz in the low and high spectral regions, respectively. In each spectral region, f_0 DLs improved dramatically with increasing f_0 as approximately the tenth harmonic appeared in the passband. Generally, f_0 DLs for complexes with similar harmonic numbers were similar in the two spectral regions. The dependence of f_0 discrimination on harmonic number presents a significant challenge to autocorrelation (AC) models of pitch, in which predictions generally depend more on spectral region than harmonic number. A modification involving a ''lag window''is proposed and tested, restricting the AC representation to a limited range of lags relative to each channel's characteristic frequency. This modified unitary pitch model was able to account for the dependence of f_0 DLs on harmonic number, although this correct behavior was not based on peripheral harmonic resolvability. [DOI: 10.1121/1.1904268]

PACS numbers: 43.66.Hg, 43.66.Ba, 43.66.Fe [NFV] Pages: 3816–3831

I. INTRODUCTION

Psychophysical studies of the pitch of harmonic tone complexes have demonstrated a relationship between the ability to discriminate small differences in fundamental frequency (f_0), and the harmonic numbers presented, i.e., the ratios between the frequencies of the individual harmonic components and f_0 of the stimulus (Houtsma and Goldstein, 1972; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). Harmonic complexes containing components with frequencies less than ten times the f_0, i.e., harmonic numbers below the tenth, generally yield good f_0 discrimination performance, while those containing only harmonics above the tenth yield poorer f_0 discrimination performance, at least for f_0's in the 100 to 200 Hz range. The different f_0 discrimination results yielded by low- and high-order harmonics have traditionally been explained in terms of harmonic resolvability (Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994). The individual frequency components of a harmonic complex are spaced linearly in frequency, while auditory filter bandwidths are approximately proportional to the filters' characteristic frequencies (CFs). The frequency spacing between low-order harmonics will be wider than the bandwidths of the auditory filters they excite. As a result, one low-order harmonic will dominate the output of a single auditory filter, and will therefore be resolved by the auditory system. Conversely, multiple high-order harmonics fall within a single auditory filter and will therefore be unresolved by the auditory system. To estimate the f_0, the individual frequencies of resolved low-order resolved components, derived from either rate-place or temporal phase-locking cues, could be compared to an internally stored harmonic template (Goldstein, 1973; Wightman, 1973; Terhardt, 1974, 1979; Srulovicz and Goldstein, 1983). A separate temporal mechanism could estimate the f_0 for unresolved harmonics, by acting on the temporal envelope resulting from the interaction of several components within a single auditory filter, which has a periodicity corresponding to the f_0 (Moore, 1977; Shackleton and Carlyon, 1994; Cariani and Delgutte, 1996a).

Certain results in the literature have provided evidence that f_0 discrimination performance is related to harmonic resolvability. One such result concerns the effect of phase on f_0 discrimination. Houtsma and Smurzynski (1990) showed that both the magnitude and the phase-dependency of f_0 difference limens (DLs) varied with harmonic number in the same way. Complexes containing low-order harmonics yielded small f_0 DLs that were not affected by the relative phase of the individual harmonics, whereas complexes containing only high-order components yielded large f_0 DLs that were phase-dependent. The phase relationship between harmonics should only affect f_0 discrimination if the harmonics are unresolved and interact within a single auditory filter (Moore, 1977; Shackleton and Carlyon, 1994). Therefore, the co-occurrence of large and phase-dependent f_0 DLs suggests that f_0 discrimination performance also depends on harmonic resolvability. Another important result concerns the

[a]Author to whom correspondence should be addressed; electronic mail: jgbern@mit.edu

ability to hear out the frequency of an individual harmonic of a complex, which is a more direct estimate of harmonic resolvability. Bernstein and Oxenham (2003) found that f_0 DLs showed the same dependence on harmonic number as listeners' abilities to hear out harmonic frequencies. Below about the tenth harmonic, f_0 DLs were small and the frequency of an individual harmonic could be heard out. Above the tenth harmonic, f_0 DLs became large, and individual component frequencies were no longer discriminable from nearby pure-tone frequencies.

These studies have shown that f_0 discrimination performance has the same dependence on harmonic number as two different measures that clearly depend on harmonic resolvability. Nevertheless, this is not conclusive evidence that f_0 discrimination is directly dependent on harmonic resolvability. In fact, several results in the literature suggest that f_0 discrimination performance does not depend on harmonic resolvability *per se*. Bernstein and Oxenham (2003) showed that the dichotic presentation of harmonic complexes, where even and odd components were presented to opposite ears, did not increase the harmonic number of the transition between good and poor f_0 discrimination, even though twice as many peripherally resolved components were available. Similar results were shown with the dichotic presentation of two-tone complexes in normal-hearing (Houtsma and Goldstein, 1972) and hearing-impaired listeners (Arehart and Burns, 1999). These results raise the possibility that the correlation between the dependencies of f_0 DLs and harmonic resolvability on harmonic number is epiphenomenal and not causal.

As an alternative to harmonic template theories, pitch could be derived from a single temporal mechanism that acts on timing information from all frequency channels, regardless of resolvability (Licklider, 1951; Meddis and Hewitt, 1991a,b; Cariani and Delgutte, 1996a; Meddis and O'Mard, 1997; de Cheveigné, 1998). A recent implementation of these timing-based models is the Meddis and O'Mard (1997) unitary autocorrelation (AC) model of pitch perception. The Meddis and O'Mard model performs an AC of the probability of firing as a function of time in each simulated auditory nerve fiber (ANF). These individual autocorrelation functions are then summed across all frequency channels to produce a summary autocorrelation function (SACF). The AC in each channel contains peaks at a period equal to the inverse of the f_0 whether it responds to the envelope of the waveform of several interacting components or to an individual resolved frequency component at a multiple of the f_0. Therefore, the SACF will contain a large peak at the inverse of the f_0, allowing the extraction of the f_0. This mathematical formulation is analogous to calculating the all-order interval histogram based on spike times in the auditory nerve (Cariani and Delgutte, 1996a). Meddis and O'Mard (1997) have argued that this AC model can account for the effect of harmonic number on f_0 discrimination indicated by the psychophysical results of Shackleton and Carlyon (1994). They claimed that the AC responds inherently differently to resolved and unresolved harmonics, yielding the requisite f_0 discrimination behavior.

However, Carlyon (1998) disputed this claim, suggesting that any deterioration in f_0 discrimination seen in the AC model is a function of the roll-off of phase locking with absolute frequency, as was seen in the physiological recordings of Cariani and Delgutte (1996a), and not a function of harmonic number as seen in psychophysical studies (Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). According to Carlyon (1998), the most important shortcoming of the Meddis and O'Mard (1997) AC model is that it fails to predict the effect of harmonic number on f_0 discrimination seen in the psychophysics: two harmonic complexes with different f_0's, bandpass filtered in the same spectral region, yield very different f_0 discrimination performance when one complex contains low-order harmonics and the other does not (Shackleton and Carlyon, 1994).

The present study addressed this controversy. The Meddis and O'Mard (1997) unitary AC model of pitch perception was tested for its ability to account for the effects of harmonic number on f_0 discrimination. A psychophysical experiment measuring f_0 DLs as a function of f_0 for fixed spectral regions was performed in order to provide more data points than the six (two f_0's times three spectral regions) tested by Shackleton and Carlyon (1994). The same stimuli were then passed through the Meddis and O'Mard (1997) AC model to determine its ability to predict the experimental results. Overall, the AC model failed to predict the experimental results. Whereas the experimental results (described in Sec. II) showed decreasing f_0 DLs with increasing f_0, the model predictions (described in Sect. III) showed the opposite trend. A number of possible modifications to the model were then tested. Of these, the most successful was one similar to that suggested by Moore (1982), in which place dependency is introduced into the model, such that each frequency channel responded only to a limited range of periodicities related to the channel's CF.

II. EXPERIMENT: F_0 DLs WITH A FIXED SPECTRAL ENVELOPE

A. Rationale

This experiment was intended to provide a larger data set than that provided by Shackleton and Carlyon (1994) with which to test the ability of the AC model to account for the effects of harmonic number on f_0 discrimination. This experiment also addressed two issues surrounding the mechanisms underlying pitch processing: the roles of phase and temporal fine-structure in f_0 discrimination.

1. Phase

Previous results have shown that the phase relationship between harmonics affects f_0 DLs (Houtsma and Smurzynski, 1990): harmonic stimuli with ''peakier'' wave forms, such as sine- or cosine-phase complexes, yield smaller f_0 DLs than those with ''flatter'' waveforms, such as random- or negative Schroeder-phase (Schroeder, 1970) complexes. However, this phase effect was not apparent in the results of Bernstein and Oxenham (2003). There are two possible reasons for this discrepancy. First, different groups of listeners were tested for the two phase relationships (random and sine phase) in the Bernstein and Oxenham (2003) study, yielding an analysis of variance (ANOVA) with less statistical power

than would be expected if the same subjects had been tested for both phase relationships. Second, as discussed by Bernstein and Oxenham (2003), listeners in their study may have performed the f_0 discrimination task without extracting the f_0, by listening for a change in the frequency of the lowest harmonic present. Although the lowest harmonic number presented was randomized from interval to interval, a large enough change in f_0 would overcome this small amount of randomization. Data analysis showed that for complexes containing only high-order harmonics, f_0 DLs were large enough that subjects may have been using the lowest harmonic cue, rather than f_0 cues, to perform the task, especially for the random-phase stimuli. If subjects were not using f_0 to perform the task, then the effects of phase on f_0 extraction would not be apparent in the results.

To address the possibility that the lack of a significant phase effect resulted from different groups of listeners participating in two phase conditions, all subjects in the present study participated in both the sine-phase and random-phase conditions. To address the possibility that listeners had used the frequency of the lowest harmonic rather than f_0 cues to perform the f_0 discrimination task, the experiment described in the following attempted to eliminate lowest harmonic cues by using harmonic stimuli with a fixed spectral envelope, and measuring f_0 DLs as a function of f_0. Although the frequency of the lowest harmonic increases with increasing f_0, a lower-numbered harmonic will also begin to appear at the low end of the spectrum. Thus, the cochlear excitation pattern will remain roughly constant, at least for those complexes containing only unresolved harmonics where the lowest harmonic cue may have played a role. As f_0 increases, the lowest harmonic number present in the passband decreases, allowing a direct comparison with the f_0 DL measurements of Bernstein and Oxenham (2003). Results indicating larger f_0 DLs in this experiment would indicate that subjects may have been using the lowest harmonic cue in the previous study.

2. Temporal fine structure

The effects of phase on f_0 discrimination have provided evidence that the pitch of complexes containing unresolved harmonics is derived from the repetition rate of peaks in the temporal envelope. Negative Schroeder-phase complexes, which have flatter envelopes than sine-phase complexes, yield larger f_0 DLs (Houtsma and Smurzynski, 1990). When unresolved harmonics are presented in alternating sine and cosine phase, yielding temporal envelopes with two peaks per period, the resulting pitch percept is judged to be twice the f_0 (Shackleton and Carlyon, 1994). Still, this does not rule out that periodicity information could be extracted from the fine structure of unresolved harmonic complexes in some conditions. Hall *et al.* (2003) demonstrated that phase manipulations affected amplitude-modulation (AM) rate discrimination performance for unresolved components in a high spectral region, but had little effect in a relatively low spectral region. Their interpretation was that fine-structure cues, which are unaffected by phase manipulations, are used in the low-frequency region, while envelope cues, which are affected by phase manipulations, are used in the high-frequency region where there is little phase-locking to the fine structure. Similarly, Bernstein and Oxenham (2003) found that for unresolved complexes containing the same harmonic numbers, f_0 DLs were larger for a 200- than a 100-Hz f_0, which may reflect reduced fine-structure information in the higher spectral region occupied by the 200-Hz complexes. Furthermore, deterioration in phase locking to the frequencies of individual partials could affect f_0 DLs for complexes containing resolved harmonics.

This experiment tested whether the presence of phase-locking to the fine-structure in the low region aided performance, in a task more closely related to pitch processing than the AM rate discrimination task of Hall *et al.* (2003). Fundamental frequency DLs were measured in two conditions: a ''low spectrum'' condition (2.5–3.5 kHz), in which phase-locking to fine structure is thought to be more available, and a ''high spectrum'' condition (5–7 kHz), in which phase-locking to the fine-structure information is greatly reduced, at least in mammalian species that have been tested so far (Rose *et al.*, 1968; Johnson, 1980; Palmer and Russell, 1986; Weiss and Rose, 1988). Testing f_0 DLs in two different frequency regions also provided a control to verify that f_0 discrimination performance depends primarily on harmonic number, and not f_0 *per se*.

B. Methods

Five subjects participated in the experiment (ages 18–21, three female). All subjects had normal hearing (15 dB HL or less re ANSI-1969 at octave frequencies between 250 Hz and 8 kHz) and were self-described amateur musicians with at least 5 years of experience singing or playing a musical instrument.

All stimuli were presented in modified uniform masking noise (UMN_m; Bernstein and Oxenham, 2003). This noise is similar to uniform masking noise (UMN; Schmidt and Zwicker, 1991), in that it is intended to yield pure-tone masked thresholds at a constant sound pressure level (SPL) across frequency, but the spectrum is somewhat different; UMN_m has a long-term spectrum level that is flat (15 dB/Hz SPL in our study) for frequencies below 600 Hz, and rolls off at 2 dB/octave above 600 Hz. The noise was low-pass filtered with a cutoff at 16 kHz. Thresholds for pure tones at 200, 500, 1500, and 4000 Hz in UMN_m in the left ear were estimated via a three-alternative forced-choice, 2-down, 1-up adaptive algorithm (Levitt, 1971). For each subject, pure tone thresholds in UMN_m fell within a 5 dB range at all four frequencies tested, such that harmonic components presented at equal SPL had nearly equal sensation level (SL). As an approximation, we defined 0 dB SL for each subject as the highest of the thresholds across the four frequencies tested, which ranged from 31 to 33.3 dB SPL across all subjects.

The stimuli were generated digitally and played out via a soundcard (LynxStudio LynxOne) with 24-bit resolution and a sampling frequency of 32 kHz. The stimuli were then passed through a programmable attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented to the subject via Sennheiser HD 580 headphones. Subjects were seated in a double-walled sound-attenuating chamber.

Fundamental frequency DLs as a function of a complex's f_0 were estimated via a three-alternative forced-choice, 2-down, 1-up adaptive algorithm tracking the 70.7% correct point (Levitt, 1971). The f_0 difference (Δf_0) was initially set to 10% of the f_0. The starting step size was 2% of the f_0, decreasing to 0.5% after the first two reversals, and then to 0.2% after the next two reversals. The f_0 DL was estimated as the average of the f_0's at the remaining six reversal points. If the standard deviation of the last six reversal points was greater than 0.8%, the data were rejected and the run repeated. In each trial, two of the three intervals contained harmonic complexes with a base $f_0(f_{0,\text{base}})$, while the other interval contained a complex with a higher $f_0(f_{0,\text{base}}+\Delta f_0)$. Subjects were informed that two of the intervals had the same pitch, while the third interval had a higher pitch, and were asked to identify the interval with the higher pitch. DLs were estimated for six different f_0's in each spectral condition (low: 50, 75, 100, 150, 200, and 300 Hz; high: 100, 150, 200, 300, 400, and 600 Hz). The f_0's tested in the high condition were double those tested in the low condition such that the harmonic numbers presented were the same in each spectral region. Measurements were repeated four times per subject for each combination of frequency region, phase, and f_0, except for one subject who completed only two runs for the random-phase conditions.

Stimuli were resynthesized for each trial of the experiment. First, diotic harmonic complexes containing equal-amplitude harmonics of the f_0 up to 10 kHz were synthesized. These harmonic complexes were then filtered with both fourth-order low-pass and fourth-order high-pass digital Butterworth filters. The 3-dB filter cutoff frequencies for the high- and low-pass filters, respectively, were 2.5 and 3.5 kHz in the low condition, and 5 and 7 kHz in the high condition. The filter weights for the high-pass filters were scaled such that the double filtering operation gave a 0-dB maximum amplitude response. The duration of the stimulus in each trial of the experiment was 500 ms, including 30-ms Hanning window onset and offset ramps.

Following the filtering operations, the stimulus in the interval with the higher f_0 was scaled in amplitude to have equal rms power to that of the two other intervals. The complexes were presented at an average level per component (before filtering) of 15 dB SL per component (adjusted individually based on tone-in-noise detection thresholds). In order to prevent the use of loudness cues, amplitude randomization was applied by roving the amplitude of the complex in each interval by ± 5 dB, uniformly distributed. On the average, the following -15 dB (re max) frequency bands contained harmonics above threshold: 1.56–5.35 kHz and 3.28–9.37 kHz in the low and high spectral conditions, respectively.

The resulting signals were then added to the UMN_m noise. Because of the rms normalization step, the average presentation level *per harmonic* was somewhat higher for the interval with $(f_{0,\text{base}}+f_0)$ then for the intervals with $f_{0,\text{base}}$. However, this difference was quite small relative to the 10 dB random amplitude variation, reaching only about 0.6 dB for the largest measured f_0 DL of 15%. Complexes were presented in either sine or random phase. For the random-phase stimuli, the phase of each harmonic was newly chosen from a uniform random distribution ranging from $-\pi$ to $+\pi$ in each interval of the experiment.

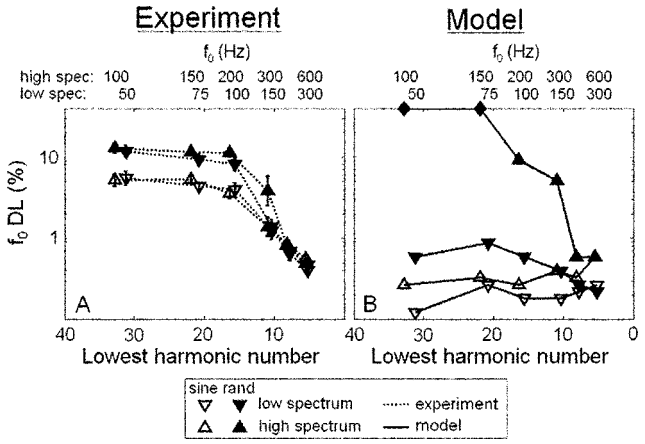


FIG. 1. Fundamental frequency DLs (a) measured psychophysically and (b) predicted by the optimal detector autocorrelation model, as a function of the lowest harmonic number present within the passband. Stimulus f_0's corresponding to the lowest harmonic numbers are listed at the top. Optimal model predictions in (b) are calculated as the minimum value of δ such that d' exceeds the value of $d'_0=190$ depicted in Fig. 3. Closed diamonds plotted along the top horizontal axis indicate that d'_0 was not reached at the maximum tested value of $\delta=0.3$.

C. Results and discussion

For each frequency region condition and f_0, the lowest detectable harmonic number (N) was estimated by dividing the average lowest detectable frequency in the passband (1.56 and 3.28 kHz in the low and high conditions, respectively) by the f_0. Figure 1(a) shows the estimated f_0 DLs as a function of N. The corresponding f_0's in the low- and high-spectrum conditions are shown along the top axis. Figure 1(b) shows the f_0 DLs predicted by the autocorrelation model, which will be discussed in Sec. III. The main findings of this experiment are (i) f_0 DLs increase with increasing N (decreasing f_0), independent of spectral region, (ii) the relative phase relationship between partials affected f_0 DLs for high, but not low N, and (iii) there was a small but significant effect of spectral region on f_0 DLs. Each of these effects will be discussed in turn.

A repeated-measures analysis of variance (RMANOVA) with three within-subject factors (spectral region, phase, and N) was conducted in order to determine the influence of each factor on f_0 discrimination. Values of $p<0.05$ were taken to be statistically significant. The RMANOVA was performed with logarithmically transformed data in an attempt to satisfy the equal-variance assumption, and the Greenhouse–Geisser (Geisser and Greeenhouse, 1958) correction for sphericity was included wherever necessary, with corrected values for degrees of freedom reported. However, neither manipulation affected the statistical significance of any main effect or interaction. Data from the subject who completed only two runs in the random-phase conditions were excluded from the RMANOVA. While the six f_0's tested in the high-spectrum

TABLE I. Results of the RMANOVA for the f_0 DL experiment. Asterisks indicate statistical significance ($p<0.05$). Degrees of freedom are adjusted based on the Geisser–Greenhouse correction.

	Effect	F	df	p
Main effects	N	161	(2.12, 6.37)	<0.0005*
	Phase	4180	(1, 3)	<0.0005*
	Spectral region	12.9	(1, 3)	0.037*
Two-way interactions	N*spectral region	0.827	(1.45, 4.35)	0.827
	N*phase	25.1	(2.37, 7.11)	<0.0005*
	Phase*spectral region	0.144	(1, 3)	0.318
Three-way interaction	N*phase*spectral region	0.226	(1.87, 5.59)	0.15

conditions were exactly double those tested in the low-spectrum condition, the low edge frequency in the high-spectrum conditions was not exactly double that of the low-spectrum conditions. As a result, N's differed by approximately 5% in the two spectral conditions. Nevertheless, for the purpose of performing the RMANOVA, we assumed that the N's were identical. For example a 100 Hz low-spectrum stimulus was assumed to have the same N as a 200 Hz high-spectrum stimulus. This small 5% shift in the value of N was unlikely to affect the RMANOVA results. The results of this analysis are shown in Table I.

There is clear transition from large to small f_0 DLs as f_0 increases (N decreases) in both the low- and high-spectrum conditions. The dependence of f_0 DLs on N is supported by a significant main effect of N. The transition to small f_0 DLs occurs as the approximately tenth harmonic [the highest resolved harmonic as estimated by Bernstein and Oxenham (2003)] begins to appear at the low end of the passband, consistent with previous results (Houtsma and Smurzynski, 1990). When plotted as a function of N, the low- and high-spectrum data overlap, indicating that f_0 DLs in these conditions depend mainly on harmonic number and not on f_0 or spectral region. This conclusion is supported by the fact that there was no significant interaction between spectral region and N.

Phase effects are apparent in these results, but only for those complexes with $N>10$, where random-phase f_0 DLs are larger than sine-phase f_0 DLs, consistent with previous findings (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003). The significant main effect of phase and a significant interaction between phase and N is consistent with the observation that phase effects are only observed for stimuli with high N. For low f_0's ($N>10$), the random-phase relationship of the harmonics gives f_0 DLs of 11%–13%, which are much poorer than had been previously measured for random-phase complexes containing only high-order harmonics (Bernstein and Oxenham, 2003). This result indicates that the previous estimates of f_0 DLs in the 6%–8% range for high-order, random-phase complexes likely reflected the influence of the "lowest harmonic present" cue (see Bernstein and Oxenham, 2003). The relatively small f_0 DLs ($\sim$4%–6%) measured for the sine-phase, high-order complexes were approximately the same as those measured in the Bernstein and Oxenham (2003) study, suggesting that the lowest harmonic cue did not play a role in the sine-phase conditions. With the elimination of the confounding "lowest harmonic" cue that affected random-phase but not sine-phase f_0 DLs, the effects of phase on f_0 discrimination are found to be significant, in line with Houtsma and Smurzynski (1990). The large f_0 DLs would make music perception based on unresolved complexes difficult, since musical semitones are only 6% apart in frequency.

While there was a significant main effect of spectral region, f_0 DLs for the same N did not generally appear to be different between the low- and high-spectrum conditions, with one exception: performance was notably worse for the high-spectrum stimulus in the random-phase, $N\approx10$ case. This difference was only observed for two of the five subjects, one of whom showed very large variability across runs, and does not constitute a general trend in the data. Although there was neither a significant two-way interaction between spectral region and either N or phase, nor a significant three-way interaction, the main effect of spectral region disappeared when the $N\approx10$ data were excluded from the RMANOVA analysis [$F(1,3)=4.8$, $p=0.12$]. This implies that phase locking to the stimulus fine structure did not play a significant role overall in f_0 discrimination for the stimuli used in this experiment.

The lack of a main effect of spectral region or a significant interaction between N and spectral region conflicts with the results of Hoekstra (1979), who also measured f_0 DLs as a function of f_0 for bandpass-filtered harmonic complexes in various spectral regions. Comparing similar spectral regions to those used in the current experiment, Hoekstra found that f_0 DLs were larger at higher spectral regions for complexes with small N, but not large N, suggesting that phase-locking to the stimulus fine-structure is more important for low-order, resolved harmonics. The discrepancy between the results of Hoekstra (1979) and the current study may be related to the bandwidths of the spectral regions used in the two studies. Hoekstra's 1/3-octave filters yielded only one audible partial for those stimuli with a low enough N to yield small f_0 DLs, while the approximately one-octave filters used in the current study produced multiple audible partials for all stimuli. The different results obtained in the two studies suggest that phase-locking to the stimulus fine structure may be more important for pure-tone frequency discrimination than for complex-tone f_0 discrimination. Alternatively, it may be that temporal fine-structure information is important for complex-tone f_0 discrimination, but that a large effect of spectral region was not observed in the present study because of the frequency ranges chosen for the two spectral conditions. The 3-dB bandpass-filter cutoff frequencies were chosen such that phase-locking should have been greatly reduced in the high-spectrum condition relative to the low-spectrum condition. However, the filter slopes yielded an audible frequency range in the high-spectrum condition that extended down to 3.28 kHz, where phase-locking to the stimulus fine structure might still have been available.

III. SIMULATIONS WITH THE AUTOCORRELATION MODEL

A. Introduction

Meddis and O'Mard (1997) showed that the autocorrelation model successfully accounted for the results of Houtsma and Smurzynski (1990): for stimuli with a fixed f_0, f_0 DLs increased as the order of the harmonics increased. Carlyon (1998) suggested that the model's successful prediction was due not to its dependence on harmonic number and harmonic resolvability, but to the reduction of phase-locking with increasing absolute frequency. Because Houtsma and Smurzynski (1990) tested only one stimulus f_0 of 200 Hz, it was not clear from their results whether the increase in f_0 DLs was due to effects of harmonic number and resolvability, or to effects of spectral region. Consistent with earlier studies (Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003), the present experiment, which measured f_0 DLs in two different spectral regions, demonstrated that f_0 discrimination performance depended mainly on harmonic number, and not spectral region or f_0. These data provide a basis for testing the Meddis and O'Mard (1997) autocorrelation model to determine its ability to predict the dependence of f_0 discrimination on harmonic number.

B. Model description

The stimuli from our psychophysical experiment were passed through the Meddis and O'Mard (1997) autocorrelation model to determine its ability to account for the psychophysical f_0 discrimination results. This model consists of an outer/middle ear bandpass filter, a basilar membrane gammatone filterbank (Patterson *et al.*, 1992), inner hair cell half-wave rectification and low-pass filtering, and the translation of the inner hair cell membrane potential into a probability of firing versus time in the auditory nerve fiber. The model used to generate ANF firing information in these simulations was identical to that used by Meddis and O'Mard (1997), except for the following two changes. First, 40 channels, consisting of only those CFs falling within the stimulus passband (1.5–5 kHz and 3–10 kHz for the low- and high-spectrum conditions, respectively) were used, with CFs spaced according to the Greenwood (1961) human scale. CFs falling outside these ranges, where the harmonic complex stimuli would not be detectable in the psychophysical experiment, were not included. Second, the inner hair-cell and auditory nerve models were replaced by a newer model (Sumner *et al.*, 2002) that allowed for stochastic spike generation. All ANFs were modeled as high spontaneous-rate fibers. The bandwidths of the model's gammatone filters were derived from the equivalent rectangular bandwidth (ERB_N) formula described by Glasberg and Moore (1990), just as in the Meddis and O'Mard study. Because the only physiologically derived cochlear mechanical filtering data available for humans (Shera *et al.*, 2002) are only appropriate for very low-level stimuli, the psychophysical bandwidths derived by Glasberg and Moore (1990) form a reasonable substitute.

Two different methods for converting from ANF firing to a psychophysical f_0 DL estimate were tested. The first method was that used by Meddis and O'Mard (1997), whereby discriminability was estimated by the Euclidean distance (D) between autocorrelation functions (ACFs) calculated from the ANF probabilities of firing as a function of time, $p(t)$. The second method was an optimal detector model based on stochastic firing of the ANFs. These methods are described in the following two sections.

1. Euclidean distance measure

Meddis and O'Mard's (1997) procedure for estimating discrimination thresholds was also used here. The main difference was that whereas Meddis and O'Mard based all of their computations on $p(t,k)$, the probability of firing (p) as a function of time (t) for each ANF channel index (k), the current simulations were based on stochastic ANF responses. This allowed for the possible influence of ANF refractoriness on the results. The inner hair cell/auditory nerve complex was set to "spike" mode (Sumner *et al.*, 2002), yielding stochastic boolean responses $s(t,k)$, whereby a one or a zero represented the presence or absence of a spike at each point in time. Each stimulus was resynthesized and presented to the model $n=15$ times (although n was increased to 60 and 100 for the simulations to be described in Secs. IV B and IV E, respectively) and $p(t,k)$ was estimated by averaging across the n outputs $s(t,k)$ obtained for each k.

The autocorrelation function (ACF) of $p(t,k)$ was then calculated in each fiber according to the formulation of Meddis and O'Mard:

$$h(t_0,l,k)=\frac{1}{\tau}\sum_{i=1}^{\infty} p(t_0-T,k)p(t_0-T-l,k)e^{-T/\tau}dt \quad (1)$$

where $h(t,l,k)$ is the channel's ACF, t_0 is the point in time at which the autocorrelation was measured, l is the autocorrelation lag, τ is the autocorrelation time constant, dt is the sampling interval, 25 μs, and $T=idt$. Because of the exponential window used in the ACF formulation, the autocorrelation will tend to fluctuate with time. In these simulations, t_0 was chosen to be an integer number of periods of each stimulus, just before the beginning of the offset ramp. This is in contrast to the Meddis and O'Mard study, where a "snapshot" of the SACF was taken at the end of the stimulus. The only other difference in the autocorrelation calculation in this study as compared to Meddis and O'Mard (1997) was that here τ was selected to be 25 ms, whereas Meddis and O'Mard used a shorter τ of 10 ms. The τ used in the current study, being longer than the period corresponding to the minimum f_0 tested, 50 Hz, tended to smooth out the SACF variation across time. A summary autocorrelation function, $SACF(f_0,l)$, was computed by summing the individual channel ACFs. The range of lags was fixed throughout the modeling from zero to a maximum lag (l_{max}) of 25 ms. This value of l_{max} corresponds to a minimum frequency of 40 Hz, which is below the minimum f_0 of 50 Hz used in our psychophysical experiment.

For each combination of f_0, spectral region, and phase, ACFs and SACFs were calculated for stimuli with f_0 increased by small perturbations, Δf_0, with 30 values of $\delta = \Delta f_0 / f_0$ logarithmically spaced across the range $0.001 \leqslant \delta$

$\leqslant 0.3$. Following Meddis and O'Mard (1997), the squared Euclidean distance between the SACFs of the unperturbed stimulus ($\delta = 0$) and each of the perturbed stimuli was then calculated:

$$D^2(f_0,\delta) = \sum_{i=0}^{l_{\max}/dt} [\mathrm{SACF}((1+\delta)f_0, idt) - \mathrm{SACF}(f_0, idt)]^2. \quad (2)$$

The procedure to convert from the D^2 statistic to an estimate of the f_0 DL was to choose a criterion based on a threshold D^2 (D_0^2), which served as a free parameter in fitting the model predictions to the psychophysical data. The lowest value of δ producing a D^2 that exceeded D_0^2 was taken to be the estimated f_0 DL. (In practice, to reduce erroneous results due to noise in the data, D^2 was judged to exceed D_0^2 only if it did so for two consecutive values of δ.) Because D_0^2 was allowed to vary as a free parameter, the D^2 measure was unable to predict an absolute f_0 DL that could be directly compared with experimental data. Rather, this statistic yielded a measure of the relative discriminability between stimulus pairs, providing a way to compare trends in the SACF and trends in measured f_0 DLs across different conditions.

2. Optimal detector model

The D^2 measure is a potentially flawed decision variable. Because D^2 is simply the distance between two SACF functions, it is likely to be sensitive to changes in stimulus dimensions that are unrelated to the stimulus pitch. For example, whereas psychophysical f_0 discrimination performance is fairly robust to changes in stimulus bandwidth, Pressnitzer *et al.* showed that such changes affect the SACF amplitude, and therefore model predictions based on the D^2 statistic. Similarly, Carlyon (1998) demonstrated that the D^2 statistic is susceptible to changes in stimulus amplitude, such as those introduced by level roving in the current study. Although calculating the D^2 between SACF functions averaged across many stimulus trials would reduce the influence of level roving on the model predictions, such a strategy would be likely to fail on a trial-by-trial basis due to its sensitivity to SACF amplitude fluctuations. An optimal detector model, with the ability to incorporate the variance associated with level roving into the decision statistic, was tested as a possible alternative.

The operation of the optimal detector was based on signal detection theory (Green and Swets, 1966). Up to four different sources of noise were present in the model: (1) the stochastic firing of the ANF; (2) stimulus level roving; (3) the background noise; and (4) phase randomization. Only the first two noise sources were always present in the simulations. For the initial simulations, background noise was not used, while phase randomization was only present in the random-phase conditions. These noise sources produced SACF variation at each lag, allowing the performance of an optimal detector to be computed based on the statistical properties of the SACF variation.

The decision variable was assumed to be a vector $\Delta\overline{\mathrm{SACF}}(f_{0A}, f_{0B})$ containing the SACF differences (ΔSACF) yielded at each lag by two stimuli with different f_0's (f_{0A} and f_{0B}):

$$\Delta\mathrm{SACF}(f_{0A}, f_{0B}, l) = \mathrm{SACF}(f_{0A}, l) - \mathrm{SACF}(f_{0B}, l). \quad (3)$$

In this model, the optimal detection strategy—the weighting of the information obtained at different lags—will vary depending on the f_0 and Δf_0. As in the D^2 model, each stimulus was presented $n = 15$ times for the each combination of f_0, frequency region, phase, and δ. Each $s(t,k)$ was substituted for $p(t,k)$ in Eq. (1) to yield stochastic individual channel ACFs, which were then summed across channels to yield n stochastic SACFs.

The performance (d') achieved by an optimal detector for discriminating stimuli on the basis of f_0 was estimated to be

$$(d')^2 = \Delta\overline{m}^T G^{-1} \Delta\overline{m}, \quad (4)$$

where $\Delta\overline{m}$ was the mean of the $\Delta\overline{\mathrm{SACF}}$s across the n stimulus trials, and G is the covariance matrix, calculated from the n $\Delta\overline{\mathrm{SACF}}$s (Van Tress, 2001). In practice, both the mean and variance of ΔSACF were nearly zero for a subset of lags, such that G was often nearly singular and not easily invertible. To resolve this problem, a very small amount of independent noise (variance $= 10^{-8}$) was added to each lag by augmenting the variances along the diagonal of G.

Because the d' estimates obtained from Eq. (4) will vary depending on the number of nerve fibers and the number of lag points used in the simulation, no attempt was made to predict the experimental d' value of 1.26 (2-up, 1-down, 3AFC, Hacker and Ratcliff, 1979) using the model simulations. The extremely large d' estimates reported in the following are a result of the large number of individual observations of f_0-related activity available across the lag range, and are not reliable estimates of absolute performance. Instead, a similar procedure to the D^2 method was used, whereby a d' criterion (d'_0) was chosen in order to predict an f_0 discrimination threshold, allowing relative performance comparisons across conditions.

C. Stimuli

The stimuli were produced in the same manner as those in our experiment, including level roving and phase randomization applied independently to each of the n stimulus presentations. There were three main differences between the stimuli used in the experiment and those used in the modeling simulations. First, the stimuli used in the modeling were reduced in duration to 200 ms in order to reduce computational load. The shorter duration should have no effect on the model predictions, since the autocorrelations were calculated only near the end of each stimulus, with a relatively short $\tau = 25$ ms and an $l_{\max}$ of 25 ms. Furthermore, decreasing the stimulus duration has little effect on f_0 DLs until durations fall below about 100 ms (Plack and Carlyon, 1995). Thus, it can be assumed that these 200-ms stimuli would yield similar results to the 500 ms stimuli used in our psychophysical experiment.

Second, no background noise was used in the initial model simulations. The main reasons for using a background noise in the psychophysical experiment (to mask distortion products and to promote the fusion of individual components into a single object) are not issues for the autocorrelation model with linear gammatone filters. However, because the presence of a background noise may still affect the ANF response to the complex tone stimuli, the possible influence of a background noise on the simulation results is examined in Sec. II E.

Third, the method of setting the signal levels differed from the psychophysical experiment. Because the model contained only high spontaneous-rate ANFs, the dynamic range available to human listeners was not available to the model. Stimulus levels similar to those actually used in the experiment tended to saturate the ANF outputs. To determine a reasonable operating level for the modeling simulations, it was assumed that for a given stimulus level, an optimal detector would choose to use those ANFs that yield the best possible performance, and discard those ANFs that yield little information, as in the "selective listening hypothesis" (Delgutte, 1982, 1987; Lai *et al.*, 1994). In these simulations, rather than adjusting the model ANF spontaneous rates and thresholds to find those that yielded the optimal performance for a given stimulus level, the ANF parameters were kept fixed and the stimulus level was adjusted. Pilot tests indicated that the best overall performance (in terms of both D^2 and d') occurred when the firing rate (r) of an ANF with CF at the center of the stimulus passband was at approximately the 90% point of the operating range, that is, when $r = r_{sp} + 0.90(r_{max} - r_{sp})$, where r_{sp} and r_{max} are the spontaneous and maximum ANF firing rates, respectively. Therefore, in the simulations all stimulus levels were set such that a pure tone at the level and frequency of a harmonic component at the center of the stimulus passband yielded an r at 90% of the operating range of an ANF with CF at the tone frequency. Athough the absolute model performance was best at this stimulus level, the relative performance of the model across the various conditions was generally unaffected by the stimulus level, provided the stimuli were above rate threshold.

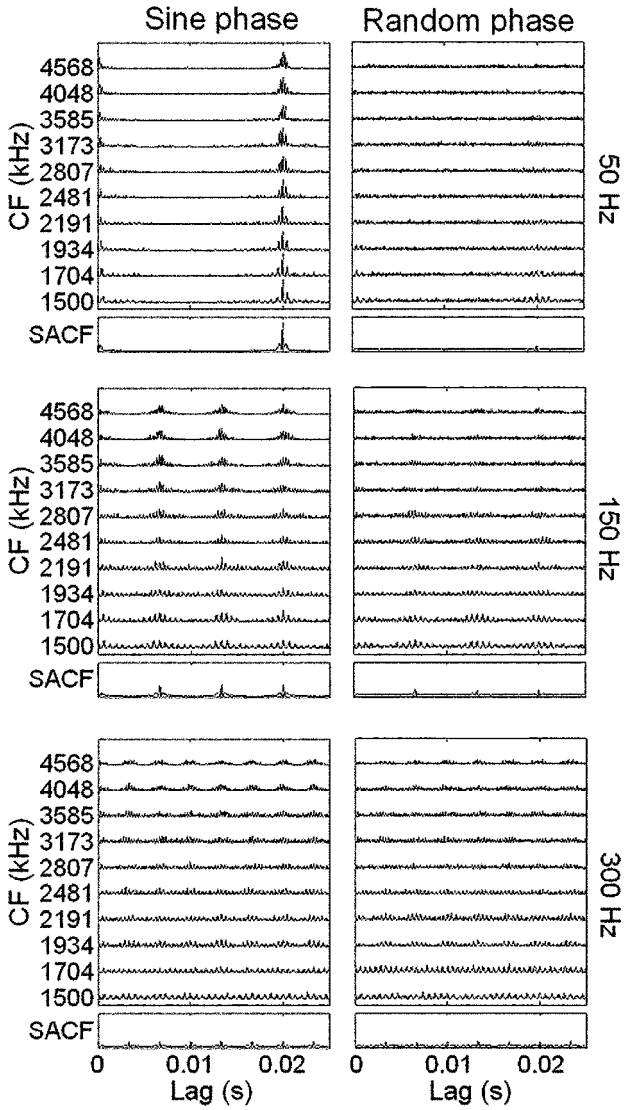


FIG. 2. Sample ACFs (top ten plots in each panel) for a subset of model ANFs with CFs as indicated along the vertical axis, for low-spectrum stimuli with three selected f_0's under both phase conditions. The corresponding SACFs are shown in the bottom plot of each panel.

D. Model results

The two main findings of the simulations are that (1) the D^2 and d' formulations of the model yield virtually identical predictions, and (2) neither formulation was successful in accounting for the psychophysical results, especially for the sine-phase conditions.

1. Comparison of the D^2 and d' measures

The Euclidean distance and optimal detector procedures produced virtually identical results. Because both procedures yield the same results, only the optimal detector model will be shown and discussed for the remainder of the paper. That these two procedures yielded similar results is perhaps not surprising, since both measures involve taking the sum of the squares of the differences between SACF functions. The main difference between the two methods is that the d' method weights these differences based on the variances at different lags across stimulus trials, whereas the D^2 statistic weights each lag equally. The similar results seen for the two methods suggests that the weighting was of little consequence—lags falling between SACF peaks added little to the sum of squared differences between SACFs, regardless of the weighting strategy. The finding implies that the D^2 measure was in fact sensitive to f_0-related activity in the SACFs, and that weighting the lags equally yields results similar to those yielded by an optimal strategy.

It is important to note that in these simulations, the Euclidean distance procedure was not challenged with level roving, which was essentially eliminated by averaging SACFs across stimulus trials. On a trial-by-trial basis, the simple Euclidean distance measure might be more sensitive to the level roving than to the changes in f_0, prohibiting it from detecting changes in f_0. In contrast, the optimal detector formulation took into account the variance due to level roving. The similarity of the two sets of results suggests that the optimal detector model was able to ignore level roving effects in discriminating f_0.

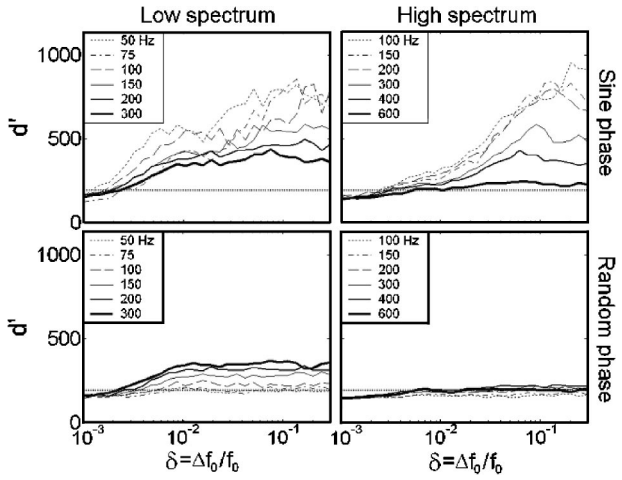


FIG. 3. Plots of the estimated d' as a function of δ, the fractional change in f_0, as predicted by an optimal detector model. For sine-phase stimuli, slopes decrease with increasing f_0, while for random-phase stimuli, slopes increase with increasing f_0. Horizontal dotted lines indicate the arbitrary $d_0'=190$ used to predict f_0 discrimination thresholds plotted in Fig. 1(b). The plots rarely cross, indicating that the predicted f_0 DL vs f_0 trend is independent of the chosen value of d_0'.

2. Optimal detector predictions

Figure 2 shows SACFs and individual channel ACFs for low-spectrum complexes with three different f_0's. Sine-phase stimulus responses are shown in the left column. For the lowest f_0 of 50 Hz, harmonics are all unresolved and interact within each model filter, such that the ACFs in each channel are phase-locked to the stimulus envelope. For the middle f_0, 150 Hz, harmonics begin to be resolved for the lowest CFs, and ACFs in these channels become phased-locked to individual sinusoids rather than stimulus envelopes. At 300 Hz, harmonic resolvability extends further, up to about 2.4 kHz. Amplitudes of SACF peaks are largest for the 50 Hz case where the f_0 appears to be coded mainly by the envelope, and diminish with increasing f_0, as resolved harmonics appear. A similar effect was observed in the high-spectrum conditions, where the SACF peaks were even smaller (not shown).

The observed decrease in SACF peak amplitude with increasing f_0 for sine-phase stimuli is reflected in the model's f_0 DL predictions. Figure 3 shows the model's predicted d' as a function of δ, the fractional change in f_0. Figure 1(b) shows the minimum values of δ such that $d'>d_0'$, where $d_0'=190$ was arbitrarily selected (horizontal dashed line in Fig. 3) to yield predicted f_0 DLs in the general range of the psychophysical results. For the sine-phase stimuli (open symbols), predicted f_0 DLs generally *increase* with increasing f_0, opposite to the trend seen in the psychophysical data. This is the case in both spectral regions. Note that this trend would occur independently of the chosen d_0', since the $d'(\delta)$ functions (Fig. 3) rarely cross. These results indicate that phase-locking to the envelope of unresolved harmonics was stronger than phase-locking to individual resolved harmonics, yielding smaller predicted f_0 DLs for lower stimulus f_0's. This result may depend on the relatively high stimulus spectral regions tested. Phase-locking to resolved components would most likely be stronger for stimuli with energy below 1.5 kHz, the frequency at which phase locking begins to roll off in the guinea pig-based model used here.

For random-phase stimuli [closed symbols in Fig. 1(b)], f_0 DLs predicted by the model tended to decrease with increasing f_0, consistent with the general trend seen in the psychophysical results. Diamonds indicate that d' failed to exceed d_0' for the largest tested value of $\delta=0.3$. The heights of the SACF peaks did not appear to change substantially with f_0 (Fig. 2, right column), suggesting that the decrease in f_0 DLs is most likely a result of the additional SACF peaks present for stimuli with larger f_0's. This correct behavior for the random-phase conditions is a result of a very large phase effect that is present mainly for low f_0's, where the predicted f_0 DLs for the same f_0 are drastically different between the two phase conditions. The presence of such a phase effect in the model (albeit much larger than that seen in the data) is consistent with previous studies that have found phase effects in the AC for harmonic complexes containing high-order harmonics, but not for those containing low-order harmonics (Patterson *et al.*, 1995; Meddis and O'Mard, 1997; Carlyon and Shamma, 2003). Since the autocorrelation operation discards relative timing information across channels, but remains sensitive to timing information within each channel, we expect the relative phase of harmonics to affect the resulting SACFs only in cases where the harmonics are unresolved by the cochlear filters, i.e., for the lowest f_0's presented.

For similar harmonic numbers present in the passband, the AC model predicts larger f_0 DLs in the high-spectrum conditions [triangles in Fig. 11(b)] than in the low-spectrum conditions (inverted triangles), suggesting an effect of spectral region in the model that was not seen in the psychophysical data. This is consistent with Carlyon's (1998) conclusion that, in contrast to the psychophysical results, the AC model is sensitive to spectral region effects, as a result of the decline in phase-locking with increasing absolute frequency.

E. Effects of added noise

The simulations described above were performed without the presence of background noise. To test the possibility that background noise could affect the model simulation results, a subset of the simulations were repeated with background noise present. In our psychophysical experiment, the background noise level was held fixed and the stimulus level set relative to the detection threshold for a pure tone in the noise. Repeating a similar strategy to determine an appropriate noise level for the modeling simulations would require a model for signal-in-noise detection based on ANF responses, which is outside the scope of this paper. Instead, we chose to examine the influence of background noise over a range of levels. The nominal signal level was the same as that used in our original simulations. The background noise levels were chosen such that the signal-to-noise ratio (SNR) ranged from -10 dB to $+\infty$ (no noise) relative to the average SNR used in the experiment ($\mathrm{SNR}_{\mathrm{expt}}$). The background noise was turned on 100 ms before, and off 100 ms after, the harmonic stimulus.

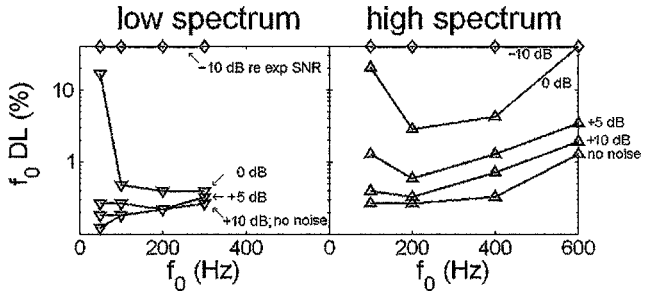


FIG. 4. Effects of the introduction of background noise on model predictions. Signal level was held constant while the noise level was adjusted; SNR (dB) are described relative to the SNR used in the psychophysical experiment. For SNRs 5 dB or greater than that used in the experiment, the background noise has little effect on model predictions. As in Fig. 1, closed diamonds plotted along the top axis indicate that d_0' was not reached for the highest δ tested of 0.3.

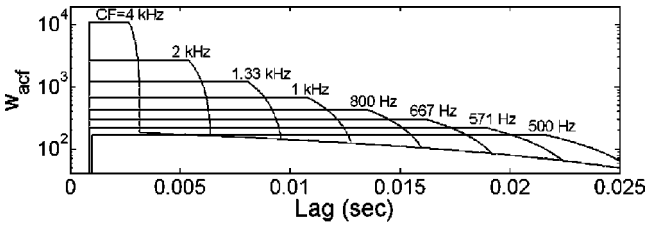


FIG. 5. Sample w_{ACF}'s [Eq. (6)] for a range of CFs, with parameters that yielded the best fit to the experimental data as shown in Fig. 7.

Figure 4 shows the predicted f_0 DLs at various SNRs (re $\mathrm{SNR}_{\mathrm{expt}}$) for the sine-phase conditions. Low-spectrum and high-spectrum results are plotted in the left and right panels, respectively. The predictions are largely unaffected by the background noise until the SNR reaches the $\mathrm{SNR}_{\mathrm{expt}}$. Interestingly, for a narrow window of SNRs near $\mathrm{SNR}_{\mathrm{expt}}$, the trend in f_0 DLs as a function of f_0 actually switches, and f_0 DLs decrease with increasing f_0 as in the experimental data. One aspect of this behavior with respect to SNR is consistent with previous psychophysical data. Hoekstra (1979) showed that f_0 DLs generally increase with decreasing SNR, and that this effect is most pronounced in a given fixed frequency region for low f_0's at low SNRs. In the model simulations, the predicted f_0 DLs increase more rapidly with decreasing SNR for low f_0's than for high f_0's. However, Hoekstra (1979) also showed that the general trend for f_0 DLs to improve with increasing f_0 for a fixed spectral region was unaffected by SNR. In contrast, the model only shows a trend for f_0 DLs to increase with f_0 for a narrow range of SNRs, and is therefore unsatisfactory as a predictor of f_0 DL data.

Overall, this analysis shows that the model predictions are relatively unaffected by the presence of background noise, provided the SNR is above a certain threshold. For the remainder of the simulations described in the following, no noise background was used.

IV. MODEL MODIFICATIONS

To account for a variety of psychophysical effects, various modifications to autocorrelation models of pitch have been suggested. These include SACF normalization (Patterson *et al.*, 1996; Yost *et al.*, 1996; Patterson *et al.*, 2000), SACF weighting functions (Pressnitzer *et al.*, 2001; Krumbholz *et al.*, 2003; Cedolin and Delgutte, 2005), a lag-dependent AC time constant (Wiegrebe, 2001) a nonlinear filterbank (Lopez-Poveda and Meddis, 2001), and a CF-dependent ACF weighting function (Moore, 1982). In the model simulations described in the following five sections, the CF-dependent weighting function was the most successful in accounting for the effect of harmonic number observed in these psychophysical results of Sec. II. Each of these possible modifications is discussed in turn.

A. SACF normalization

The height of the SACF peak normalized to the value at zero lag has been successful in predicting the pitch strength of iterated rippled noises (Patterson *et al.*, 1996; Yost *et al.*, 1996; Patterson *et al.*, 2000). Cariani and Delgutte (1996a,b); performed an analysis similar to SACF normalization by using the peak-to-background ratio in the all-order interval histogram as a neural estimate of the pitch salience. They were able to successfully account for a wide range of psychophysical pitch phenomena using this type of analysis. However, when the optimal detector model was adjusted to include SACF normalization (results not shown), there was virtually no change from the results seen in Fig. 5. The reason for this is that the optimal detector inherently normalizes the SACF function to the standard deviation at each lag. In essence, the extra normalization step scales the mean and standard deviation of the SACF equally, leaving d' unaffected. SACF normalization did serve to reduce the noise associated with level roving, increasing the overall d'. However, this effect was similar across all conditions, such that when d_0' was adjusted accordingly, normalized and unnormalized SACFs yielded virtually identical f_0 DL predictions.

B. SACF weighting function

An SACF weighting function that generally gives more weight to short lags should yield a larger estimated d' for high-f_0 stimuli that contain SACF peaks at short lags. Thus, such a weighting function may have the potential to account for the better discrimination performance observed for high f_0's. For example, Pressnitzer *et al.* (2001) found that the Meddis and O'Mard (1997) model, modified to include a linear SACF weighting function, successfully predicted an increase in the lowest f_0 that could convey melody for higher spectral regions. In the optimal detector formulation, weighting the SACF would have no effect, since the weights would alter both the mean and standard deviations by the same factor, thus not affecting d'. Instead, independent noise with variance $\sigma_w^2(l)$ was added along the diagonal of the covariance matrix G in Eq. (4), according to

$$\sigma_w^2(l) = w(l)^{-2}, \tag{5}$$

where $w(l)$ is the analogous SACF weighting function. Three different versions of $w(l)$ were tested: a linear function, $w=1-l/l_{\max}$ (Pressnitzer *et al.*, 2001), a power function, $w=1-(l/l_{\max})^{\alpha}$ with α ranging from 1/64 to 1 (Krumbholz *et al.*, 2003), and an exponential function, $w=\exp(-l/\lambda)$ with λ ranging from 0.3 to 30 (Cedolin and Delgutte, 2005). For each w, the model was tested both with and without SACF normalization. The most promising results were produced by the combination of an exponential $w(l)$ with $3<\lambda<4$ ms, and SACF normalization. For low-spectrum stimuli, this modified model yielded f_0 DLs that decreased with increasing f_0 for low-spectrum stimuli, consistent with the experimental data (results not shown). However, this combination of modifications was unable to account for the high-spectrum data, and was therefore unsatisfactory. None of the other functions produced desirable results.

C. A lag-dependent time constant

Another lag-dependent AC modification was suggested by Wiegrebe (2001), whereby the AC time constant [τ in Eq. (1)] increases with increasing lag. Like the SACF weighting function of Pressnitzer *et al.* (2001), a lag-dependent τ would affect the SACF differently for different stimulus f_0's, and could therefore influence the model's f_0 DL predictions. However, this modification would most likely not account for the results of the experiment described in Sec. II, because the longer time constant associated with low f_0's would tend to increase the amplitudes of peaks in the SACF, yielding smaller f_0 DLs than for high f_0's. Thus, Wiegrebe's (2001) modification would be likely to skew the model predictions even more heavily in favor of low f_0's.

D. A nonlinear filterbank

The described above model simulations used a bank of linear gammatone filters (Patterson *et al.*, 1992) to represent the basilar membrane. A more accurate nonlinear filter model that includes the compressive input–output function observed at the level of the basilar membrane (Rhode, 1971; Ruggero *et al.*, 1997) has been shown to be important for a number of psychophysical phenomena (e.g. Oxenham and Bacon, 2003), and might better account for the f_0 DL data. The inclusion of a basilar membrane nonlinearity (e.g., Lopez-Poveda and Meddis, 2001) might compress the "peaky" sine-phase waveform more than the "flat" random-phase waveform yielded by interacting unresolved harmonics (Carlyon and Datta, 1997), possibly reducing the size of the phase effect predicted by the AC model. However, simulations using the dual-resonance nonlinear (DRNL) filterbank (Lopez-Poveda and Meddis, 2001) yielded unsatisfactory results (not shown), similar to those seen with the gammatone model. Thus, although the compression offered by this model is similar to that observed physiologically, it was not substantial enough to account for these data.

E. A CF-dependent "lag window"

Section III showed that for sine-phase stimuli, the Meddis and O'Mard (1997) AC model responded preferentially to low f_0's for stimuli bandpass filtered in fixed spectral regions. Therefore, to successfully predict the improved f_0 discrimination for higher f_0's seen in the human performance, the AC model must be modified in such a way as to impair performance for low f_0's within a given spectral region. One way to accomplish this is to limit the range of lags for which the autocorrelation is calculated in each frequency channel in a CF-dependent manner (Moore, 1982). With this lag-window limitation, the AC will respond best to f_0's that have certain harmonic numbers falling within each channel's bandwidth. Schouten (1970) first proposed the idea that "each pitch extractor has a limited range of measurable time intervals" in order to account for Ritsma's (1967) demonstration of the dominance of low-order harmonics in complex pitch perception. Moore (1982) further quantified the lag window, suggesting that a mechanism based on first-order interspike intervals operates over a range of lags between about 0.5/CF and 15/CF. Thus the AC in a particular channel will respond to f_0's that are 1/15 to 2 times the channel's CF. In other words, the AC will respond to a given f_0 only if at least one of the f_0's first to fifteenth harmonics fall near the CF. Ghitza (1986) implemented a similar idea, whereby the interspike interval analysis window length was roughly inversely proportional to each channel's CF.

After experimenting with various possibilities, we found that a piecewise-linear weighting function was able to account for the psychophysical data with some success. The CF-dependent weighting function consisted of four segments:

$$w_{\mathrm{ACF}}(l,CF)=\begin{cases}0, & l<\dfrac{0.5}{CF}\\[2ex] \dfrac{CF^2}{CF_0}, & \dfrac{0.5}{CF}\leqslant l<\dfrac{N_C}{CF}\\[2ex] \dfrac{CF^2}{CF_0}-m\left(l-\dfrac{N_C}{CF}\right), & \dfrac{N_C}{\mathrm{CF}}\leqslant l<\dfrac{(N_C+N_\Delta)}{CF}\\[2ex] A-\dfrac{A}{l_0}l, & l\geqslant\dfrac{(N_C+N_\Delta)}{CF},\end{cases} \tag{6}$$

where l is the lag, $\mathrm{CF}_0=1500$ Hz, the lowest CF used in the simulations, N_C is the cutoff between the second and third segments relative to CF, N_Δ is the width of the third segment relative to CF, A is the amplitude of the fourth segment at $l=0$, l_0 is the lag for which the fourth segment reaches zero, and m, the slope of the third segment, is defined as

$$m=\frac{\mathrm{CF}^2-A+\dfrac{A}{l_0}\left(\dfrac{N_C+N_\Delta}{\mathrm{CF}}\right)}{N_\Delta/\mathrm{CF}}. \tag{7}$$

The fourth segment, independent of CF, is identical to the linear SACF weighting function of Pressnitzer *et al.* (2001). The zero-crossing of this segment (l_0) was set to 33 ms as suggested by Pressnitzer *et al.*, consistent with a 30 Hz lower limit of melodic pitch. Finally, in some conditions, the estimated d' reflected activity at low lags completely unrelated to the stimulus f_0. To prevent this phenomenon, w_{ACF} for

each CF was set to zero for all values of $l<0.875$ ms. Sample w_{ACF} functions for various CF are shown in Fig. 5. (The linear segments of the functions appear curved because they are plotted on a logarithmic scale.) The lag window was applied to the ACF for each simulated ANF, and these windowed ACFs were summed to create the SACF just as before.

The CF-dependent windowing procedure described here was notably different from the SACF weighting described in Sec. IV B. There, the addition of independent noise was used as a substitute for a SACF-weighting function, which would have scaled the mean and standard deviation equally, yielding no net effect on d'. Here, the weighting functions (w_{ACF}) were applied to the individual ACFs before summing them to produce the SACF. Thus, the statistical properties of the SACF at each lag tended to reflect the statistical properties of the ACFs for channels that were most heavily weighted at that lag.

Estimates of d' were generally noisier than in the unmodified model because the lag window tended to reduce the total number of ANF spikes that were used in the calculation. Therefore, two minor modifications were made. First, the number of stimulus repetitions n was increased to 100. Second, d' was determined to exceed threshold only if it did not fall below d_0' again for a higher value of δ. This ensured that the threshold was not exceeded due to random fluctuations in the d' estimates.

The modified AC model was fit to the sine-phase experimental data of Sec. II with four free parameters (N_C, N_Δ, A, and d_0'). The two most important aspects of the experimental data were the dependence of f_0 DLs on N, and the lack of an effect of spectral region on f_0 DLs. Therefore, the fitting procedure minimized the sum of two error measures: the root-mean-squared difference between the logarithms of predicted and actual f_0 DLs, and the root-mean-squared difference between the logarithms of the predicted f_0 DLs for stimuli with equivalent N's in the low- and high-spectrum conditions. The strong model nonlinearities and limited range of δ values tested prohibited the successful use of an automated fitting procedure, such as the Nelder–Mead simplex method used by MATLAB's fminsearch function. Instead, a parameter-space search method was used, where coarse step-sizes allowed for a reduction in computation time. Thus, we caution that a somewhat different set of parameters may yield a better fit than those reported here.

Figure 6 shows d' as a function of δ for the modified model with parameters that yielded the best fit to the sine-phase experimental data: $N_C=10.8$, $N_\Delta=2$ and $A=200$. The sample w_{ACF} functions depicted in Fig. 5 reflect these parameter values. The best-fitting d_0' of 7.91×10^4 is depicted as a horizontal dashed line in each panel of Fig. 6. Figure 7(b) shows the modified model's f_0 DL predictions as a function of N, based on these best-fitting parameters. The f_0's corresponding to the N's are shown along the top axis. The psychophysical results from Fig. 1(a) are replotted in Fig. 7(a) for direct comparison with the model predictions. The modified model yielded a reasonable fit to both sets of data, and captured three main features of the data. First, f_0 DLs generally decrease with increasing f_0. Second, the model predictions for the two spectral regions overlap when plotted as a function of N, such that f_0 DL are mainly dependent on harmonic number. The separation of stimuli into two groups based on N is clearly seen in Fig. 6: those stimuli with low f_0's, such that $N>12$, have shallow d' vs δ slopes, yielding large f_0 DLs, while those with high f_0's, such that $N<12$, have steeper slopes, yielding small f_0 DLs. Third, phase effects are only present for complexes with large N. For small N, sine- and random-phase stimuli yield similar f_0 DL predictions.

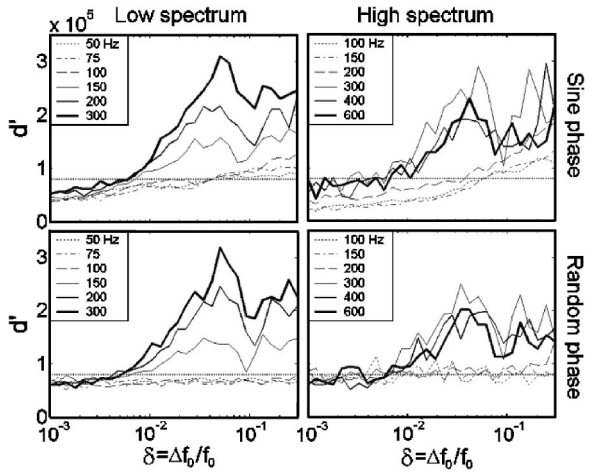


FIG. 6. Model estimates of d' vs δ using the lag windows described in Eq. (6) and pictured in Fig. 5, with parameters $N_C=10.8$, $N_\Delta=2$, and $A=200$ that best fit the sine-phase data. Stimulus f_0's are clearly divided into two groups, with lower f_0's yielding gradual d' slopes, and higher f_0's yielding steeper d' slopes.

The one major failure of the modified model is that it overpredicted the phase effect for low f_0's. The variability in the envelopes associated with low-f_0, unresolved, random-phase complexes was so large relative to the mean

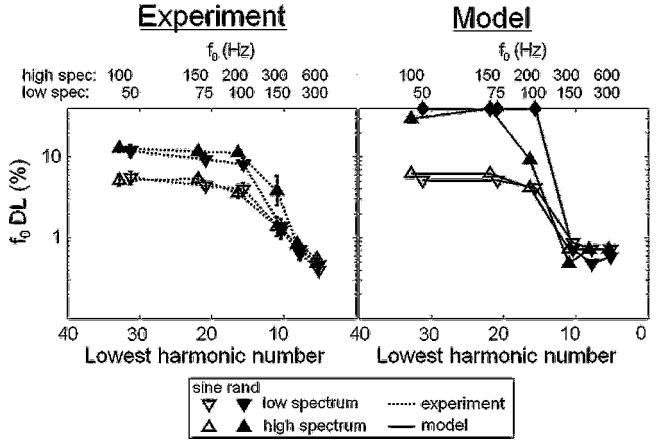


FIG. 7. (a) Psychophysical f_0 DLs from Fig. 1(a) are replotted for direct comparison with the model predictions. (b) Model f_0 DL predictions based on d' estimates shown in Fig. 6 using the lag window [Eq. (6)], plotted as a function of N. As in Fig. 1, f_0's corresponding to values of N for the low- and high-spectrum conditions are plotted above each panel, and closed diamonds plotted along the top horizontal axis indicate that d_0' was not reached at the maximum tested value of $\delta=0.3$. Both experimental and model f_0 DLs generally overlap for stimuli with the same N, indicating the modified model successfully accounts for effects of N on f_0 discrimination performance.

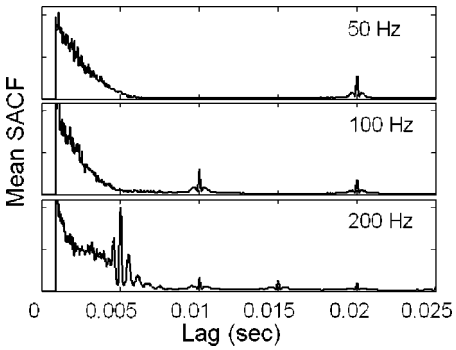


FIG. 8. Mean SACFs produced with the lag-window modification [Eq. (6); Fig. 5] for sine-phase stimuli with various f_0's. For higher f_0's (e.g., 200 Hz), the large SACF peak at $l=5$ ms contains large fine-structure side peaks, causing the nonmonotonic behavior of d' observed in Fig. 6. For lower f_0's (50 and 100 Hz), the SACF side peaks are small relative to the central SACF peak; nonmonotonic d' behavior is not observed for these stimuli.

envelope that d' was not affected by increasing δ. Thus, the model failed to reach threshold at the highest tested value of $\delta=0.3$, and was unable to predict discrimination thresholds for these complexes. This problem was also observed for the original, unmodified model. The inclusion of a compressive nonlinearity in the model might help to reduce the magnitude of this phase effect by compressing ''peaky'' sine-phase envelope more than ''flat'' random-phase envelopes. However, because substituting DRNL filters (Lopez-Poveda and Meddis, 2001) for gammatone filters did not greatly affect the predictions of the unmodified model (Sec. IV D), it is also unlikely to greatly influence the predictions of the modified model.

Nonmonotonicities were observed in d' estimates at the three highest f_0's tested in each condition. For values of δ near 0.1, d' estimates suddenly decreased then increased. This nonmonotonic behavior can be understood by examining the sample SACF functions in Fig. 8. For the relatively high f_0 of 200 Hz, the SACF contains multiple sharp peaks at lags near $1/f_0$, reflecting the stimulus fine structure. As δ increases, these closely spaced peaks move in and out of alignment with one another, yielding the observed nonmonotonic behavior. In contrast, the SACF representations for low f_0's (e.g., 50 and 100 Hz) are dominated by a single large peak at each multiple of $1/f_0$, with relatively small side bands. As a result, nonmonotonic behavior is not observed for these stimuli. This analysis suggests that the model uses fine-structure information to discriminate f_0 for low-order, but not for high-order harmonics. Regardless, these nonmonotonicities occur for f_0 separations well above the discrimination threshold, and therefore do not impact the model's f_0 DL predictions.

V. DISCUSSION

The analysis of Sec. III showed that the Meddis and O'Mard unitary AC model of pitch perception is unable to account for the dependence of f_0 DLs on harmonic number. Whereas experimental data presented both here and elsewhere (Hoekstra, 1979; Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001) show that discrimination performance deteriorates with increasing lowest harmonic number within a given passband, the Meddis and O'Mard model predicts just the opposite for the stimuli used here. This result is consistent with the results of Cedolin and Delgutte (2005), who estimated pitch salience based on all-order interval analysis of cat ANF spikes. They found that pitch salience estimated in this way was maximal for the lowest f_0's tested, where individual harmonics are not well resolved by the cat's auditory periphery.

We have shown (Sec. IV) that this failure of the AC model is not fatal to the idea that a single mechanism based on temporal information can account for the perceived pitch based on both resolved and unresolved harmonics. With the introduction of a CF-dependent lag window similar to that described by Moore (1982), the model was able to predict the dependence of f_0 discrimination on harmonic number. This was achieved because the modification reverses the original model's ''preference'' for high-order harmonics by applying a weighting function that amplifies the AC response to low-order harmonics, and attenuates the response to high-order harmonics.

The success of the modified AC model where the original model has failed supports the idea that temporal information alone is not enough to yield a salient pitch percept, and that the temporal information must be presented at the correct place on the cochlear partition in order to yield good f_0 discrimination performance (Oxenham *et al.*, 2004). The lag window modification effectively codes ''place'' information into the AC model by weighting each channel's contribution based on its relationship to the stimulus f_0. For a given CF, a range of lags between 0.5/CF and N_C/CF are weighted most heavily. The ACF will respond most readily to a certain range of stimulus f_0's that contain peaks falling within this lag range.

It is important to note that the correct behavior of the modified model with regard to the effects of harmonic number is *not* based on harmonic resolvability. The modified model responds preferentially to complexes containing low harmonics because of the introduction of the CF-dependent w_{ACF}. This could be considered a major failing of the model, if good f_0 discrimination performance were directly dependent on the presence of resolved harmonics. On the other hand, the direct dependence of the modified model's f_0 DL predictions on harmonic number is consistent the results of several studies (described in Sec. I) suggesting that f_0 discrimination performance may depend only on harmonic number, and not on harmonic resolvability *per se* (Houtsma and Goldstein, 1972; Arehart and Burns, 1999; Bernstein and Oxenham, 2003).

The AC model was modified to fit the f_0 discrimination data described in Sec. II, and has not yet been tested on other data sets. Nevertheless, the dependence of predicted f_0 performance on harmonic number is a direct result of the w_{ACF} modification, suggesting that the modified model should be able to account at least qualitatively for the results of other studies that have shown an increase in f_0 DLs with increasing N. These include f_0 discrimination studies with bandpass-filtered harmonic complexes (Houtsma and Gold-

stein, 1972; Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003), as well as those that manipulate N for complexes with a fixed f_0 (Houtsma and Goldstein, 1972; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). Furthermore, because this modification relies on harmonic number rather than peripheral resolvability, it is likely to account for results indicating that the diotic presentation of alternating harmonics does not improve f_0 discrimination performance (Houtsma and Goldstein, 1972; Arehart and Burns, 1999; Bernstein and Oxenham, 2003) despite the improvement in peripheral resolvability (Bernstein and Oxenham, 2003).

In contrast to the behavior of the modified model with respect to N, its correct behavior with respect to phase effects is most likely based on harmonic resolvability. The original model predicted a large effect of phase on f_0 DLs for low f_0's containing unresolved harmonics, where the SACF mainly reflects phase-locking to the envelope (Fig. 2). In these conditions, the envelope resulting from the interaction of multiple harmonics within one filter was much ''peakier'' with sine-phase complexes than with random-phase complexes, yielding smaller predicted f_0 DLs. While the modified model predicts that stimuli yielding large f_0 DLs should also yield phase-dependent f_0 DLs, the two effects rely on different processes. The dependency of f_0 DLs on harmonic number derives from the w_{ACF} modification. The dependency on phase derives from inherent differences in the way the model processes resolved and unresolved harmonics, and is correctly predicted by both the original and modified AC models.

How is the mathematical formulation of a lag window to be interpreted in terms of physiological mechanisms? Licklider (1951) formulated an AC model of pitch perception in terms of a system of neurons, where every cochlear frequency channel is associated with its own bank of AC neurons, and each neuron in the bank is tuned to one of a wide range of periodicities. The ACF [Eq. (1)] represents the responses of each of the neurons in the bank, and the lag window is a weighting function applied to these responses. In the physiological interpretation, a larger number of neurons associated with a given lag will reduce the noise in the periodicity representation, yielding smaller predicted f_0 DLs for the f_0 associated with that lag.

In a manner similar to that described in the harmonic template model of Shamma and Klein (2000), the autocorrelation mechanism might develop over time to detect only those temporal correlations that tend to occur in the outputs of individual ANFs in response to generic wideband stimuli. The CF-dependent lag windows described here [Eq. (6) and Fig. 5] could emerge naturally based on the statistical properties of ANF outputs in response to such stimuli. Since the temporal extent of the impulse response of a bandpass filter is inversely proportional to the filter's bandwidth, the narrower filters associated with lower CFs will yield a wider range of lags over which a filtered wideband input stimulus will correlate with itself. Mirroring the properties of these naturally occurring autocorrelations, the system would be tuned to detect ANF response correlations at longer lags for low CFs than for high CFs. De Cheveigné and Pressnitzer (2005) have proposed a similar idea that relates filter impulse response times to pitch processing.

With the addition of a CF-dependent lag window, a single pitch mechanism based on temporal information can account for the poorer f_0 discrimination performance associated with high N. However, it does not address other evidence relating to frequency modulation (FM) detection and temporal integration that points to the possible existence of two separate pitch mechanisms. Plack and Carlyon (1995) showed that f_0 discrimination was affected by decreasing stimulus durations below 100 ms more for unresolved than for resolved complexes. They suggested that the exceptionally poor FM detection performance (relative to the f_0 DL) measured for unresolved complexes resulted from an absence of the longer integration time needed to extract the f_0. Because the modified autocorrelation model needs the same integration time for a given f_0 (i.e., somewhat longer than a single pitch period, in order to yield an SACF peak at $l = 1/f_0$) regardless of resolvability, it is not likely to account for this result. It may be possible to interpret the CF-dependent weighting function as a manifestation of two pitch mechanisms. In this interpretation, the second segment of the lag-window [Eq. (6)] corresponds to the mechanism for low-order resolved harmonics, the CF-independent fourth segment represents the more poorly performing mechanism for high-order, unresolved harmonics and the third segment represents the transition between the two.

The autocorrelation model outlined here and elsewhere (e.g., Meddis and Hewitt, 1991a, b; Cariani and Delgutte, 1996a, b; Meddis and O'Mard, 1997) takes into account all-order intervals between ANF spikes. Kaernbach and Demany (1998) challenged the view that the f_0 detection mechanism takes into account anything but first-order interspike intervals. They showed that a click-train with f_0 information in its first-order interspike interval statistics was easier to discriminate from a random click train than a click train containing f_0 information in its second- and higher-order interval statistics, even though the waveform autocorrelation showed a similar peak at a lag corresponding to the f_0 in both cases. However, Pressnitzer *et al.* (2002) showed that an all-order autocorrelation based on simulated ANF responses, rather than the raw waveform, may be able to account for this phenomenon, as a result of the auditory filtering and neural transduction present in the model.

VI. SUMMARY AND CONCLUSIONS

Measurements of f_0 DLs for bandpass-filtered harmonic stimuli demonstrated that f_0 discrimination performance depends largely on harmonic number: as the ratio of a complex's f_0 to the frequency of its lowest component increases, f_0 discrimination improves. The Meddis and O'Mard (1997) unitary AC model of pitch perception fails to predict this effect of harmonic number on f_0 discrimination. While psychophysical measurements show an improvement in f_0 discrimination with increasing f_0 for bandpass filtered harmonic stimuli, the AC model predicts the opposite behavior, at least for sine-phase complexes. In order for the model to correctly predict the psychophysical results, an *ad hoc* modification

was made, whereby the lags for which the AC was measured in each frequency channel were weighted in a CF-dependent manner. This yielded f_0 DL predictions that decreased with increasing f_0, and depended mainly on harmonic number, consistent with the data. This modification works by forcing the model to respond preferentially to low numbered harmonics. The correct behavior of the model in no way reflects a preference for resolved harmonics, *per se*. Instead, the model introduces a dependence on harmonic number, without regard to harmonic resolvability.

In conclusion, this study has shown that a single autocorrelation mechanism, modified to include CF dependency, is sufficient to account for the dependence of f_0 DLs on harmonic number. Consequently, two pitch mechanisms may not be needed to explain this effect. Nevertheless, the modified autocorrelation model may not account for other evidence for two pitch mechanisms, such as the differences observed between resolved and unresolved harmonics in the temporal integration of f_0 information (Plack and Carlyon, 1995).

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC 05216 and by NIH Training Grant No. T32 DC 00038. We thank Christophe Micheyl, Ray Goldsworthy, Ray Meddis, Peter Cariani, Lutz Wiegrebe, and Bob Carlyon for their helpful comments on previous versions of this manuscript. We also thank Ray Meddis and Lowel O'Mard for providing the modeling software (available at http://www.essex.ac.uk/psychology/hearinglab/dsam/index.htm) and training J.G.W.B. in its use, and Bertrand Delgutte for his suggestion that the lag windows could emerge naturally based on the impulse responses of cochlear filters.

Arehart, K. H., and Burns, E. M. (**1999**). ''A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss,'' J. Acoust. Soc. Am. **106**, 993–997.

Bernstein, J. G., and Oxenham, A. J. (**2003**). ''Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?,'' J. Acoust. Soc. Am. **113**, 3323–3334.

Cariani, P. A., and Delgutte, B. (**1996a**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Cariani, P. A., and Delgutte, B. (**1996b**). ''Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch,'' J. Neurophysiol. **76**, 1717–1734

Carlyon, R. P. (**1998**). ''Comments on 'A unitary model of pitch perception' [J. Acoust. Soc. Am. 102, 1811–1820, (1997)],'' J. Acoust. Soc. Am. **104**, 1118–1121.

Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Carlyon, R. P., and Shackleton, T. M. (**1994**). ''Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?,'' J. Acoust. Soc. Am. **95**, 3541–3554.

Carlyon, R. P., and Shamma, S. (**2003**). ''An account of monaural phase sensitivity,'' J. Acoust. Soc. Am. **114**, 333–348.

Cedolin, L., and Delgutte, B. (**2005**). ''Pitch of complex tones: Rate-place and interspike-interval representations in the auditory nerve,'' *J. Neurophysiol.* (in press).

de Cheveigné, A. (**1998**). ''Cancellation model of pitch perception,'' J. Acoust. Soc. Am. **103**, 1261–1271.

de Cheveigné, A., and Pressnitzer, D. (**2005**). ''The case of the missing delay lines: Cross-channel phase interaction,'' *J. Acoust. Soc. Am.* (in press).

Delgutte, B. (**1982**). ''Some correlates of phonetic distinctions at the level of the auditory nerve,'' in *The Representation of Speech in the Peripheral Auditory System*, edited by Granstrom RCaB (Elsevier, Amsterdam), pp. 131–150.

Delgutte, B. (**1987**). ''Peripheral auditory processing of speech information: Implications from a physiological study of intensity discrimination,'' in *The Psychophysics of Speech Perception*, edited by M.E.H. Schouten (Nijhoff, Dordrecht, The Netherlands), pp. 333–353.

Geisser, S., and Greeenhouse, S. W. (**1958**). ''An extension on Box's results on the use of the *F* distribution in mulivariate analysis,'' Ann. Math. Stat. **29**, 885–891.

Ghitza, O. (**1986**). ''Auditory nerve representation as a front-end for speech recognition in a noisy environment,'' Comput. Speech Lang. **1**, 109–130.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Greenwood, D. D. (**1961**). ''Critical bandwidth and the frequency coordinates of the basilar membrane,'' J. Acoust. Soc. Am. **33**, 1344–1356.

Hacker, M. J., and Ratcliff, R. (**1979**). ''A revised table of *d′* for M-alternative forced choice,'' Percept. Psychophys. **26**, 168–170.

Hall, J. W., Buss, E., and Grose, J. H. (**2003**). ''Modulation rate discrimination for unresolved components: Temporal cues related to fine structure and envelope,'' J. Acoust. Soc. Am. **113**, 986–993.

Hoekstra, A., ''Frequency discrimination and frequency analysis in hearing,'' Ph.D. thesis, Institute of Audiology, University Hospital, Groningen, The Netherlands, 1979.

Houtsma, A. J. M., and Goldstein, J. L. (**1972**). ''The central origin of the pitch of pure tones: Evidence from musical interval recognition,'' J. Acoust. Soc. Am. **51**, 520–529.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.

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

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

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

Krumbholz, K., Patterson, R. D., Nobbe, A., and Fastl, H. (**2003**). ''Microsecond temporal resolution in monaural hearing without spectral cues?,'' J. Acoust. Soc. Am. **113**, 2790–2800.

Lai, Y. C., Winslow, R. L., and Sachs, M. B. (**1994**). ''A model of selective processing of auditory-nerve inputs by stellate cells of the antero-verntral cochlear nucleus,'' J. Comput. Neurosci. **1**, 167–194.

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

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3118.

Meddis, R., and Hewitt, M. (**1991a**). ''Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. (**1991b**). ''Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. II. Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J. (**1977**). ''Effects of relative phase of the components on the pitch of three-component complex tones,'' in *Psychophysics and Physiology of Hearing*, edited by E.F. Evans, and J.P. Wilson (Academic, London), pp. 349–358.

Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).

Oxenham, A. J., and Bacon, S. P. (**2003**). ''Cochlear compression: Perceptual measures and implications for normal and impaired hearing,'' Ear Hear. **24**, 352–366.
Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). ''Correct tonotopic representation is necessary for complex pitch perception,'' Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.
Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair cells,'' Hear. Res. **24**, 1–15.
Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). ''The relative strength of the tone and noise components in iterated rippled noise,'' J. Acoust. Soc. Am. **107**, 1578–1588.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).
Patterson, R. D., Yost, W. A., Handel, S., and Datta, A. J. (**2000**). ''The perceptual tone/noise ratio of merged iterated rippled noises,'' J. Acoust. Soc. Am. **107**, 1578–1588.
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., de Cheveigné, A., and Winter, I. M. (**2002**). ''Perceptual pitch shift for sounds with similar waveform autocorrelation,'' Acoust. Res. Lett. Online **3**, 1–6.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). ''The lower limit of melodic pitch,'' J. Acoust. Soc. Am. **109**, 2074–2084.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1968**). ''Patterns of activity in single auditory nerve fibres of the squirrel monkey,'' in *Hearing Mechanisms in Vertebrates*, edited by A.V.S. de Reuck, and J. Knight (Churchill, London).
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Schmidt, S., and Zwicker, E. (**1991**). ''The effect of masker spectral asymmetry on overshoot in simultaneous masking,'' J. Acoust. Soc. Am. **89**, 1324–1330.
Schouten, J. F. (**1970**). ''The residue revisited,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G.F. Smoorenburg (Sijthoff, Lieden, The Netherlands), pp. 41–54.
Schroeder, M. R. (**1970**). ''Synthesis of low peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.
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.
Shamma, S., and Klein, D. (**2000**). ''The case of the missing pitch templates: How harmonic templates emerge in the early auditory system,'' J. Acoust. Soc. Am. **107**, 2631–2644.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Srulovicz, P., and Goldstein, J. L. (**1983**). ''A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.
Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). ''A revised model of the inner-hair cell and auditory-nerve complex,'' J. Acoust. Soc. Am. **111**, 2178–2188.
Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.
Terhardt, E. (**1979**). ''Calculating virtual pitch,'' Hear. Res. **1**, 155–182.
Van Tress, H. L. (**2001**). *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York).
Weiss, T. F., and Rose, C. (**1988**). ''A comparison of synchronization filters in different auditory receptor organs,'' Hear. Res. **33**, 175–180.
Wiegrebe, L. (**2001**). ''Searching for the time constant of neural pitch extraction,'' J. Acoust. Soc. Am. **109**, 1082–1091.
Wightman, F. L. (**1973**). ''The pattern-transformation model of pitch,'' J. Acoust. Soc. Am. **54**, 407–416.
Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.

Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences

Francis Rumsey, Slawomir Zielinski, and Rafael Kassier
Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

Søren Bech
Bang and Olufsen, Peter Bangsvej 15, DK-7600 Struer, Denmark

(Received 18 October 2004; revised 14 March 2005; accepted 15 March 2005)

The preferences of a large number of naïve listeners were elicited in response to a selection of multichannel audio items that had been degraded in quality by using band-limiting and down-mixing algorithms. Relationships were sought between these preference ratings and the quality judgements of experienced listeners in an attempt to determine whether one could be predicted from the other. Results suggest that a simple regression model can be used to do this with adequate results, but that a better prediction can be successfully based on experienced listener ratings of timbral and spatial fidelity. There is a difference between naïve and experienced listeners in the weightings of the fidelities and their relationship to overall quality. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1904305]

PACS numbers: 43.66.Lj, 43.38.Md, 43.38.Vk [NX] Pages: 3832–3840

I. INTRODUCTION

A previous study (see Zielinski *et al.*, 2005) examined the basic audio quality of multichannel audio stimuli that were degraded by band limitation and down mixing. The results obtained using experienced listeners showed that basic audio quality is more influenced by timbral fidelity than by spatial fidelity. In order to examine whether this was as a consequence of the listening experience of the subjects, a new experiment using naïve subjects was conducted and is presented here. The prediction of the naïve subject or consumer preference ratings using the experienced listener ratings was also attempted.

It was hypothesized that the experience of the listeners used in the previous experiment would have led to them having certain prejudices and prior assumptions concerning the meaning of audio quality. In that experiment the experienced listeners were asked to grade basic audio quality, defined as ''any and all detectable differences between a reference and the item under evaluation'' (Zielinski *et al.*, 2005). Experience suggests that such basic audio quality ratings can be strongly influenced by the preference of the listener (their degree of liking), but this has never been proved in formal experiments. The scales used for evaluating basic audio quality tend to use affective terms such as excellent, bad and so forth, which are not only judgements of the difference between the reference and the evaluated item. This is even more obvious in the case of impairment scales used for judging basic audio quality, which use terms such as annoying along with perceptible (e.g., perceptible but not annoying), thereby conflating two different concepts into one scale (the one affective, the other a psychophysical judgement). Furthermore, basic audio quality ratings are often called mean opinion scores and opinion is just one step away from preference in semantic terms.

In the new experiment reported here, large numbers of so-called naïve listeners (essentially students recruited off the street with no required previous experience of evaluating music or audio) gave preference ratings of some or all of the same stimuli used in the experienced listener tests. These ratings were entirely concerned with their liking for a particular version of the audio items presented. The subjects were naïve to the extent that they had not been biased by traditional studio or academic training in audio engineering, and in this respect may be considered to be similar to ordinary consumers of audio-visual entertainment. If consumer preference could be predicted on the basis of expert listener judgements then one could avoid the need for lengthy consumer tests and rely rather on the ratings of a relatively small number of experts, using statistical models to derive predictions of the broader population's preference.

In sensory evaluations of food it is common practice to use separate panels of consumers to evaluate product preference, only using experts to evaluate detailed sensory characteristics of items. The point of using two independent panels is to enable external preference mapping to be carried out, there being the explicit assumption that the preference of experts may be biased and not representative of that of consumers. Olive, however, recently showed that, because expert ratings of audio quality are more consistent than—yet similar to—general consumers' ratings, expert ratings may be used in place of those of general consumers for loudspeaker preference tests (Olive, 2003). He argues that the evaluation of audio quality may be different to the evaluation of food because in the case of audio there is such a thing as absolute fidelity, to which all sound reproduction should aspire and which acts as a natural anchor for listeners of all sorts. It was hoped that the study reported here would help to confirm or refute this notion.

II. EXPERIMENT A

Two experiments are described in this paper (A and B). The purpose of Experiment A was to undertake a pilot study

 0001-4966/2005/117(6)/3832/9/$22.50

before carrying out the main experiment (B). There were 40 naïve listeners involved in Experiment A and another 40 naïve listeners involved in Experiment B. Subjects were recruited from the population of undergraduate students at the University of Surrey. The only selection criterion was that they were not students of sound recording, and hence could be expected to have no experience in critical listening tests. The experienced listeners used in the previously reported experiment (Zielinski *et al.*, 2005) also came from a population of university students, so it can be said that the naïve listeners are a representative group of nontrained subjects in a similar age and demographic population. (Although the subjects in both experiments came from a relatively narrow population, this is justified by the aim to compare the results for trained and untrained subjects in the same population.)

A. Stimuli

Twelve multichannel audio excerpts in the 5.0-channel format, according to ITU-R BS.775, were selected. They represented the following genres: classical music, pop music, movie, sport, TV show, and ambient sounds (applause and rain). The rationale for selection of audio material was described in more detail in Zielinski *et al.* (2003a, 2003b, 2003c, 2005). It was important that the selection of audio excerpts had a suitable spatial characteristic. For example, half of the selected items contained an F–B (foreground–background) spatial characteristic and the other half contained an F–F (foreground–foreground) spatial characteristic in order to preserve a balance in the selection of program material. In the case of F–B recordings front channels reproduce predominantly foreground audio content (mainly close and clearly perceived audio sources), whereas rear channels contain only background audio content (room response, reverberant sounds, unclear, ''foggy''). This situation may be compared to the typical sound impression perceived by a listener sitting in a concert hall (sound stage with musicians at the front, reflections from side and rear). In the case of the recordings exhibiting the F–F spatial characteristic both front and rear channels contain predominantly foreground content. This category has similarities to the auditory impression encountered when a listener is surrounded by an orchestra. Rear channels contain clearly identifiable sound sources, often different from the instruments reproduced by front channels, for example percussion instruments, backing vocals, etc. [See Zielinski *et al.* (2003a, 2003b, 2003c) and Rumsey (2002) for a detailed discussion on the categorization of audio program material according to spatial characteristics based on a scene-based paradigm.] The average duration of the selected excerpts was 20 seconds.

The quality of the stimuli was degraded using two types of signal processing: bandwidth limitation and down mixing. 11 algorithms were used for limitation of bandwidth and four algorithms for the down-mixing procedure. The rationale for selection of these algorithms and a description of them is given in detail in Zielinski *et al.* (2003a, 2003b, 2003c). Several assumptions were made before processing the audio material. First, it was assumed that the full bandwidth of any audio channel ranges up to 20 kHz. Therefore the overall bandwidth of the five-channel uncompressed audio material was assumed to be equal to 100 kHz (5×20 kHz). Second, it was assumed that in this experiment the minimum bandwidth of any low-pass filtered channel should be no less than 3.5 kHz (approximately the upper limit of telephone bandwidth). The effect of these algorithms on the overall bandwidth of the audio signal was designed to be comparable between the bandwidth-limited versions and the down-mixed versions. For example, down mixing the number of audio channels from five to three has the same effect on information rate as reducing the bandwidth of all five channels from 20 kHz to 12 kHz. A range of values was used, extending from the equivalent of five full bandwidth channels, representing the original unprocessed recordings, down to one effective channel. These spanned the whole range of audio quality (ranging from bad to excellent) and excited a broad range of perceptual impressions regarding the employed fidelity attributes. For example, the low quality anchors (3.5 kHz low-pass filtered items and mono down-mixed items) excited perceptual impressions that were on average graded using the bottom of the timbral fidelity and the spatial fidelity scales. In contrast, the top of the fidelity scales was used in order to evaluate the unprocessed items (hidden reference).

After processing, the total set of stimuli to be evaluated consisted of 138 items (including unprocessed recordings).

B. Listening test method

Tests in which naïve listeners are employed in the listening panel frequently involve simple tasks (see, for example, Zacharov and Koivuniemi, 2001b, which involved the comparison of two stimuli in terms of their preference). For Experiment A, a paired comparison test method was adopted as the basis for experimental design. The listeners were asked to listen to each pair of sounds and to rate which sound they would prefer to listen to at home, assuming that they had a choice between a number of different audio services to use with a home cinema system.

The experiments were carried out in an ITU-R BS.1116-compliant listening room using 5.0-channel surround sound reproduction. Custom-written computer software was used to automate the listening test. Buttons were used to switch synchronously between the pairs of stimuli, and sliders were provided to allow the listeners to express which stimuli they preferred. (The letters assigned to the stimuli on the user interface were chosen to correspond to adjacent keys on the computer keyboard so that they could be switched easily by that means.) For simplicity purposes it was decided to use sliders with five discrete positions. The left-hand end of the scale corresponded to a strong preference for the first stimulus whereas the right-hand end represented a strong preference for the second stimulus (the middle position was neutral).

The visual content of each stimulus was displayed on a 42 in. plasma display in front of the subjects. One listener took part in the experiment at a time. The audio quality was evaluated at the optimum listening position. Details concerning the apparatus and audio calibration procedure, including loudness equalization, are described in Zielinski *et al.* (2003c).

As there were 138 stimuli in total to be evaluated it would have been too time consuming to undertake a full paired comparison of all the stimuli, so it was decided to limit the set of compared stimuli in Experiment A. Specifically, it was decided to compare only stimuli having the same overall bandwidth, as defined above. For example, it was decided to compare low-pass filtered items, all of whose five channels were band-limited down to 12 kHz, with 3/0 down-mixed items (the overall bandwidth in these two cases is the same and equivalent to three full bandwidth audio channels). Similarly, it was decided to compare the low-pass filtered items all of whose five channels were limited down to 8 kHz with the stereo down-mixed items (overall bandwidth equivalent to two full bandwidth channels).

The rationale for these comparisons is as follows: in a previous study it was found that experienced listeners generally graded the quality of the full-bandwidth down-mixed items higher than all-channel band-limited items for a given overall bandwidth (Zielinski *et al.*, 2003a). This was caused by the fact that the changes in timbral fidelity affect quality ratings of experienced listeners more than changes in spatial fidelity (Zielinski *et al.*, 2005). However, it was still unknown whether the same results would be obtained if naïve listeners were involved in the experiment (it was hypothesized that these results could be different, perhaps because of their lack of training, expectation and absence of detailed familiarity with the characteristics of two channel stereo reproduction).

C. Results from Experiment A

The results obtained in Experiment A exhibited either binomial distribution or close to normal distribution. For this reason it was difficult to summarize and to interpret the results based on histograms. Despite the fact that the raw data was not always normally distributed, for the sake of generalizability it was decided to summarize them using mean values and 95% confidence intervals (see Fig. 1). The use of mean values and associated 95% confidence intervals can be justified in this case by the fact that the employed preference scale can be assumed to be of equal intervals, and also due to the central limit effect [the number of observations for each displayed mean value was equal to 40, which is a sufficiently large number of observations for the sampling distribution of the mean to approach the normal distribution regardless the distribution of the raw data (see, for example, Diamantopoulos and Schlegelmilch, 1997)].

Since some mean values and associated 95% confidence intervals in Fig. 1 are placed on the right-hand side of the (zero) base line, whereas some others are located on the left-hand side, it can be seen that the results of the preference evaluation are program dependent. For example, in the case of the item containing the ambient sound of rain (Ambient F–F Rain) naïve listeners preferred the five-channel low-passed filtered version (12 kHz) over the full bandwidth down-mixed one (3/0). This can be explained by the fact that in this case all channels contained almost equal energy of the sound of the rain causing, in the opinion of the authors, a very enveloping effect. The limitation in number of reproduction channels, arising from down mixing, inevitably

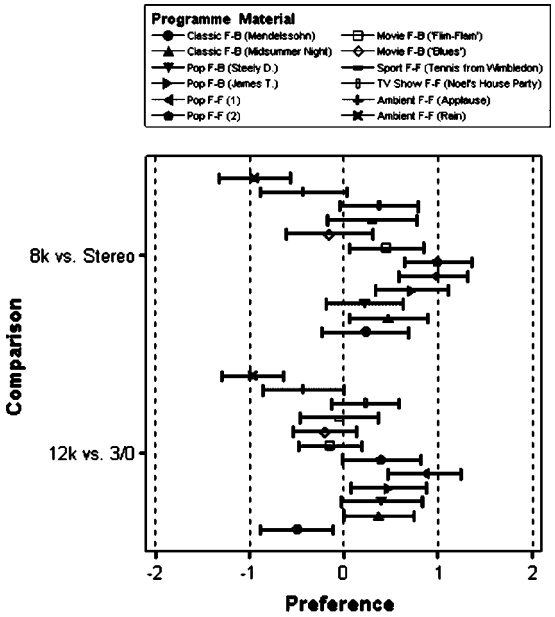


FIG. 1. Results obtained in Experiment A (means and 95% confidence intervals). Scores averaged across listeners. (On the *x*-axis, 2=strongly prefer down-mix and −2=strongly prefer low-pass filtered version.)

caused a substantial reduction in spatial fidelity, which, for the naïve listeners, was more objectionable than the degradation of the timbral fidelity caused by low-pass filtering. The opposite preference was observed in some of the pop music items (see Pop F–F 2, for example), where the effect of down mixing to stereo was less objectionable than the effect of low-pass filtering, which is in line with the results obtained for basic audio quality using the experienced listeners (Zielinski *et al.*, 2003a).

D. Discussion

The results obtained in this pilot experiment can to some extent be described as contradictory to the results obtained using a group of trained listeners (Zielinski *et al.*, 2003a). It was found that the preferences of naïve listeners depend strongly upon the program material presented. For some recordings, naïve listeners prefer limitation in bandwidth over limitation in the number of channels for a given overall bandwidth, whereas for other recordings they exhibit the opposite preference. Limiting the number of channels using down-mixing algorithms was usually graded higher by the experienced listeners than limiting the bandwidth of individual channels for a given overall bandwidth (Zielinski *et al.*, 2003a). The experienced listeners never graded the quality of the down-mixed items lower than the quality of the low-pass filtered items for a given overall information rate of program material. Naïve listeners, on the other hand, seemed to prefer certain low-pass filtered items. However, it is necessary to mention that the observations described above must be treated with caution since, as mentioned earlier, the raw data obtained in the experiment sometimes had binomial distribution, indicating the possible existence of two subpopulations amongst the listeners.

One can question whether the departure of these results from the results obtained in the previous experiment involv-

ing experienced listeners was due to a genuine effect caused by a different type of listener or whether it was due to a bias caused by adopting a different test paradigm (or due to a combination of both factors). At this stage it was not possible to estimate to what extent using a paired comparison test in Experiment A, as opposed to a multiple comparison test in the previous experiment, biased the obtained results. Therefore, it was decided to reduce any potential bias in the main experiment (Experiment B) by making the listening test method as similar as possible to the one employed in the experiment involving experienced listeners. The details of all modifications introduced to the test paradigm are provided below.

III. EXPERIMENT B

The initial results obtained in Experiment A indicate that the preference scores obtained from the naïve listeners can show a different trend from the basic audio quality scores obtained from the experienced listeners. However, as discussed above, those results have to be treated with some caution since one cannot exclude the possibility of bias due to the test paradigm employed. Those two factors encouraged the authors to undertake a further investigation using an improved test method reducing the aforementioned bias. The purpose of this new experiment was to examine what differences (or similarities) exist between the scores obtained from the naïve and the experienced listeners and also to see what has the greatest effect on the preference of naïve listeners: timbral fidelity or spatial fidelity. The main research questions can be summarized as follows:

(1) How similar are basic audio quality scores obtained from experienced listeners to the preference scores obtained from naïve listeners?
(2) Is it possible to use basic audio quality scores obtained from experienced listeners to predict the preferences of naïve listeners?
(3) Is spatial fidelity less important for naïve listeners than timbral fidelity?

A. Listening test method

The design of Experiment B was inspired by the work of Zacharov and Koivuniemi (2001a, 2001b). The preference mapping procedure using in their Audio Descriptive Analysis and Mapping (ADAM) was adapted for the purposes of this experiment. The applied method is essentially based on a preference mapping of direct attribute ratings obtained from experienced listeners to the preferences grades provided by naïve listeners. As a result a multivariate regression model of subjective preferences is derived. This model can be used subsequently to predict the preference ratings of naïve listeners on the basis of scores obtained from experienced listeners.

In Experiment B a few modifications were made in the listening test method in comparison with Experiment A. It was decided to use a multiple comparison method instead of the paired comparison approach. The main reason for this modification, as mentioned above, was the need for a reduction of a potential bias caused by differences in test methods. Another reason for this modification is the efficiency of a multiple comparison method when a large number of stimuli are to be evaluated (multiple comparison methods are generally less time consuming). The listeners were asked to compare several audio items (called services in the instructions for the listeners) and then to state whether they agreed with the following statement: ''I would like to listen to this audio service using my home cinema system.'' (A number of possible context-defining scenarios were considered for presentation to listeners so as to provide them with an ecologically meaningful scenario. Amongst this age group the home cinema experience was chosen as the context most likely to be recognized or imagined. A short scene setting story was provided to the subjects to place them within an appropriate ecological context.) They were required to record their answers using the sliders on a continuous Likert scale containing five labels: 50 (strongly agree), 25 (agree), 0 (neither agree nor disagree), −25 (disagree), −50 (strongly disagree). All 138 audio-visual items were included in Experiment B. During each listening session two low-quality anchors were incorporated as items to be evaluated: a low-pass filtered version of the original recording with a cutoff frequency of 3.5 kHz [as recommended in (ITU-R, 2001), which could be considered as a timbral low anchor] and a down-mixed center-channel mono version of the original recording (which could be considered as a spatial low anchor). The original (unprocessed) recordings were included among the items to be evaluated, however the listeners were not told which they were in order not to bias the results. It was felt that explicit presentation of the reference recording, as recommended in (ITU-R, 2001), might indicate to the listeners items that they should prefer, which was not the intention of the experimenters (one cannot exclude the possibility that listeners may prefer some processed items more than unprocessed ones).

40 ''new'' naïve subjects took part in Experiment B. Like the 40 subjects who took part in Experiment A, they were also recruited from the general student population at the University of Surrey.

Although it is known that presence of simultaneous video presentation has only a small effect on the audio quality grades given by experienced listeners (Zielinski *et al.*, 2003b), it was not known if this would be the case with naïve listeners. It was hypothesized that naïve listeners might be more sensitive towards the presentation of accompanying video than experienced listeners. Therefore, in order to investigate this audio-visual interaction in more detail, it was decided that half of the listening panel would undertake the experiment without accompanying video, whereas the other half of the listeners would undertake the experiment with accompanying video. Subjects were randomly divided into the two groups with and without accompanying video playback.

B. Results from Experiment B

1. Initial data analysis

The database of subjective responses acquired in Experiment B consisted of almost 7000 entries. The data was ini-

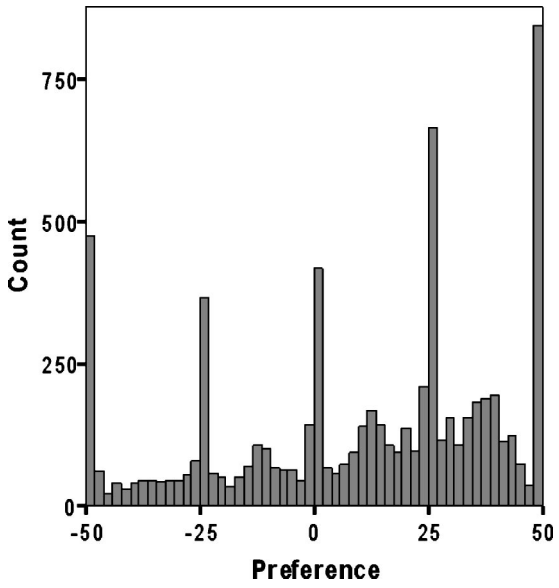


FIG. 2. Histogram of all the scores obtained in Experiment B (data from all experimental conditions).

tially examined using a histogram depicted in Fig. 2. It can be seen that the listeners used five discrete values from the scale most frequently (five peaks in the histogram). These values correspond to the five main categories, 50 (strongly agree), 25 (agree), 0 (neither agree nor disagree), −25 (disagree), −50 (strongly disagree). The obtained result shows that the gathered data is predominantly categorical in its nature and therefore care must be taken during the statistical analysis, since there is a risk of violating the assumptions underlying some statistical methods.

2. Audio-visual interaction

As mentioned in the description of the experimental design, 20 listeners undertook the experiment with the video monitor switched off, whereas another 20 listeners undertook the experiment with the video monitor switched on. This allowed the authors to check to what extent video presence/absence affected the audio preferences of the naïve listeners. In order to investigate this interaction in more detail it was decided to undertake an analysis of variance test (ANOVA). The ANOVA model consisted of three fixed factors, audio algorithm (16 levels), audio recording (12 levels), picture (two levels, On/Off), and all interactions between them. It was checked that the main assumptions underlying the ANOVA test were fulfilled. For example, the distribution of the residuals was examined using the Kolmogorov–Smirnov test and found to be normal. The assumption of homogeneity of variance was tested using Levene's test. The obtained results showed that variances were not equal across experimental conditions. However, an ANOVA test may still give reliable results even when variances are not equal across different groups, provided that the number of cases in each group is similar (Howell, 1997). This condition was not fulfilled in the experiment since some stimuli were evaluated more frequently than others (unbalanced design). A check was therefore made using a balanced version of the data set, and the results of the ANOVA test were found to be very similar to those reported below, suggesting that the slight unbalancing of the data set did not have a marked effect on the outcome. Finally, it can be assumed that the obtained scores were independent due to the experimental design (random order of presentation of stimuli for each listener).

According to the obtained results the picture factor was statistically significant [$F(1,6686)=5.07$, $p=0.024$], however according to the partial eta squared value it had only a small effect on audio preference of the naïve listeners ($\eta^2=0.001$). A significant interaction between the program material and picture presence was also observed [$F(15,6686)=3.41$, $p<0.001$], however according to the partial eta squared value the magnitude of this interaction was small ($\eta^2=0.008$). The remaining interactions involving the picture factor were statistically insignificant ($p>0.05$). The obtained results are in line with the results presented in (Zielinski *et al.*, 2003b; Beerends and De Caluwe, 1999).

In summary, the results obtained in this section show that the effect of picture presence on evaluation of audio quality undertaken by the inexperienced listeners is small.

3. Preprocessing of data

In order to compare the results obtained in the previous experiment involving experienced listeners (Zielinski *et al.*, 2003a) with the results of this experiment, it was decided to aggregate scores across listeners (including both the with picture and without picture conditions) prior to further analysis. This operation helped in mapping the results between the two experiments and also helped in the exploration and visual inspection of the data (it is common practice to reduce the size of the database by averaging the scores across the subjects in order to increase readability of the results and to help to explore the data (see, for example, Esbensen, 2002)). After this operation the size of the database was reduced from 6962 to 138 entries. [It is worth noting that after the data reduction all important information regarding the mean values for all investigated experimental conditions was preserved, the only information that was discarded was the data describing the variability between and within subjects. The operation of the data reduction also affected its distribution. Due to the central limit effect all the peaks that were present in an initial histogram (Fig. 2) were smoothed and consequently the overall histogram became less peaky.]

The scores obtained for the unprocessed items (hidden reference) were removed from the analysis. This was done in order to reduce the bias caused by the difference in experimental design between the experiment involving the experienced listeners (Zielinski *et al.*, 2003a) and this experiment. In the former experiment the listeners knew that the hidden reference was included as one of the evaluated items and they were expected to grade it using the maximum value of the scale. On the contrary, in this experiment the listeners did not know that the unprocessed items were included as items to be evaluated and consequently they were free to use any values from the scale in their evaluation.

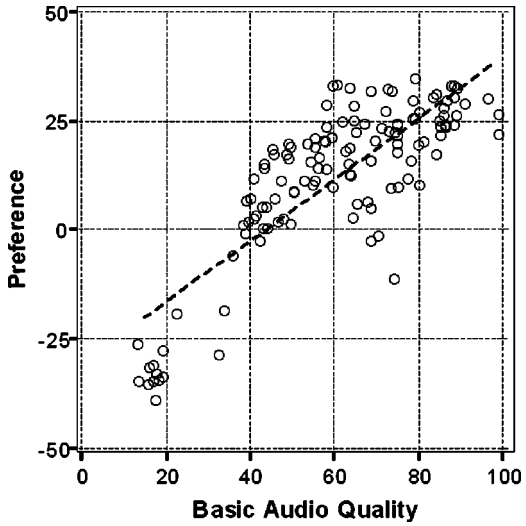


FIG. 3. Preferences of naïve listeners plotted against the basic audio quality scores obtained from experienced listeners (scores averaged across listeners). The regression model is represented using the dashed line.

4. Preference versus basic audio quality

The first research question in this study was: ''How similar are the basic audio quality scores obtained from experienced listeners to the preference scores obtained from naïve listeners?'' In order to answer this question preferences grades obtained from the naïve listeners were plotted as a function of the basic audio quality scores obtained from the experienced listeners (see Fig. 3). As can be seen in the plot, the scores are correlated with each other. According to a Pearson test, the correlation between these scores is high and equals 0.82. This result was also confirmed by Spearman's test which showed that the correlation between the scores equals 0.75.

The second research question stated before the experiment was: ''Is it possible to use the basic audio quality scores obtained from the experienced listeners to predict the preferences of the naïve listeners?'' In order to answer this question linear regression analysis was performed on the data. The summary of the obtained model is presented in Table I whereas its coefficients are given in Table II. The obtained results show that the preferences of the naïve listeners can be predicted, to some degree, using the following equation (involving unstandardized coefficients):

$$\mathrm{Pref} = 0.701\ \ \mathrm{BAQ} - 30.5, \tag{1}$$

where Pref is the naïve listener preference rating and BAQ is the basic audio quality scores obtained from experienced listeners.

The regression line representing the model is depicted in the previously discussed Fig. 3. According to the obtained results, the regression model explains 67% of the variance in the preferences of the naïve listeners. The error of prediction expressed as the standard deviation of the residuals is equal to 10.4 points relative to the 100-point scale, which is acceptable in the view of the fact that the listening tests often involve an inherent error of the order of 5% or more (it was checked that the confidence intervals of the scores obtained from the naïve listeners in this experiment ranged from 4 to 8 points relative to the 100-point scale). Since the values of R^2 and adjusted R^2 statistics are of similar magnitude, it is possible to conclude that the model cross-validates well. Consequently the obtained model, although not ideal, can be used for a crude prediction of the preference scores.

TABLE I. Basic regression model (summary).

Correlation R	R^2	Adjusted R^2	Std. Dev. of residuals
0.82	0.670	0.667	10.4

TABLE II. Basic regression model coefficients.

	Unstandardized coefficients		Standardized coefficients		Significance
	B	Std. error	Beta	t	p
(Constant)	−30.522	2.825		−10.803	0.000
Basic audio quality	0.701	0.044	0.819	15.872	0.000

As mentioned above in Sec. II B 1, care must be taken when undertaking statistical analysis due to the risk of violation of the assumptions underlying some of the statistical methods (this risk is caused by the predominantly categorical nature of the data). There is an ongoing debate as to whether it is legitimate to analyze categorical data using parametric methods, e.g., linear regression or analysis of variance. Nevertheless, some statisticians claim that it is legitimate to use a categorical predictor in the regression provided the outcome variable (basic audio quality in our case) is continuous (Field, 2000). Although it was found that the basic audio quality scores also exhibit some peaks in the histogram, their magnitude was much smaller than those observed in the case of the preference scores of naïve listeners. If it can be assumed that the basic audio quality variable is a continuous one then the approach adopted in this paper is legitimate. However, as an additional precaution, the distribution of the residuals obtained in the linear regression was scrutinized. According to the obtained results the distribution of the residuals was normal. Consequently, it can be concluded that the requirement for a normal distribution of residuals was fulfilled. In addition, taking into account that the order of stimuli assigned to the buttons in the graphical user interface was randomized for each listener, it can be assumed that the requirement for the independence of scores was also fulfilled.

The results presented in this section can be summarized as follows:

(i) There is a relatively large similarity between the basic audio quality scores acquired from the experienced listeners and the preference scores elicited from the naïve listeners (this answers the first research question).

(ii) The preference scores obtained from the naïve listeners can be, to some extent, approximated using solely the basic audio quality scores elicited from the experienced listeners.

TABLE III. Regression model summary.

Correlation R	R^2	Adjusted R^2	Std. Dev. of residuals
0.86	0.736	0.729	9.3

(iii) The obtained regression model is not very accurate; nevertheless, it can be useful if one wants to roughly predict the preferences of naïve listeners based on the results of the basic audio quality scores acquired from experienced listeners (this answers the second research question).

5. Preference versus direct attributes

It was found that a slightly more accurate regression model could be derived when three direct attributes judged by the experienced listeners were used as predictors (the term direct attributes is used to indicate attributes that are more specific than the basic audio quality; no claim is made that these attributes are orthogonal and unidimensional). The main descriptors of the obtained regression model are summarized in Table III whereas its coefficients and associated statistical significance levels p are presented in Table IV. A scatter plot of the relationship is shown in Fig. 4. As can be see in Table IV, the regression model in this case is expressed using the following equation (using unstandardized coefficients):

$$\text{Pref} = 0.633\ \text{Timbral} + 0.249\ \text{Surround} - 44.8, \quad (2)$$

where Pref is the preference scores of the naïve listeners, Timbral is the timbral fidelity judged by the experienced listeners, Frontal is the frontal spatial fidelity judged by the experienced listeners, and Surround is the surround spatial fidelity judged by the experienced listeners.

The detailed definitions of the above attributes are presented in (Zielinski *et al.*, 2005). According to the obtained regression model, the preference scores acquired from the naïve listeners depend on the timbral fidelity (0.633) and, to a lesser extent, on the surround spatial fidelity (0.249). The frontal spatial fidelity, according to the t-test, does not contribute to the model in a statistically significant way ($p > 0.05$) and therefore was omitted in the equation. Thus, it can be concluded that the preferences of the naïve listeners are primarily influenced by the timbral fidelity and, to a smaller extent, by the surround spatial fidelity, whereas the frontal spatial fidelity has no significant effect on their preferences. This leads to the important conclusion that for the naïve listeners timbral fidelity is a more important predictor of preference than spatial fidelity. In this respect the naïve listeners are similar to the experienced listeners who base their basic audio quality evaluation more on the timbral fidelity rather than on the spatial fidelities.

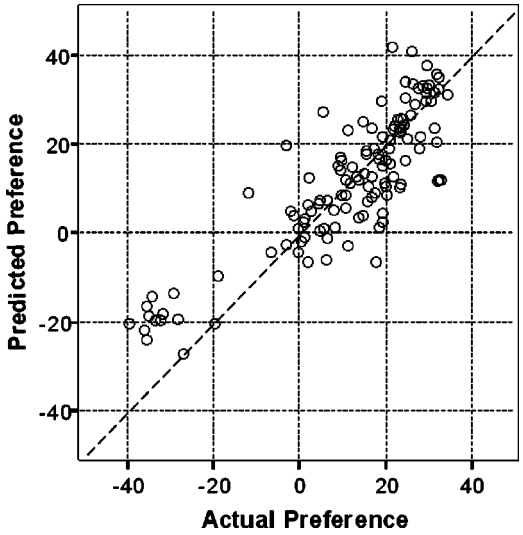


FIG. 4. Scatter plot of predicted versus actual preferences of naïve listeners using three direct attributes from experienced listeners. Dashed lines represents the diagonal $y = x$.

6. Similarities and differences between experienced and naïve listeners

As already demonstrated in Sec. II B 4, the basic audio quality scores obtained from the experienced listeners are correlated with the preferences of naïve listeners. Moreover, it was shown that the naïve listeners, similarly to the experienced listeners, pay more attention to timbral changes than to spatial changes when evaluating surround sound reproduction.

The main difference between the experienced and the naïve listeners can be demonstrated using Fig. 5. This figure shows the regression coefficients from the two models obtained using the experienced and naïve listeners, respectively. The first model (labeled as Exp) allows one to predict the basic audio quality scores based on the scores given to the three direct attributes and was derived in our previous experiment (Zielinski *et al.*, 2005). This model is expressed by the following equation (using unstandardised coefficients):

TABLE IV. Regression model coefficients.

	Unstandardized coefficients		Standardized coefficients		Significance
	B	Std. error	Beta	t	p
(Constant)	−44.796	3.661		−12.237	0.000
Timbral fidelity	0.633	0.039	0.785	16.400	0.000
Frontal spatial fidelity	0.000	0.053	0.000	−0.008	0.994
Surround spatial fidelity	0.249	0.044	0.331	5.672	0.000

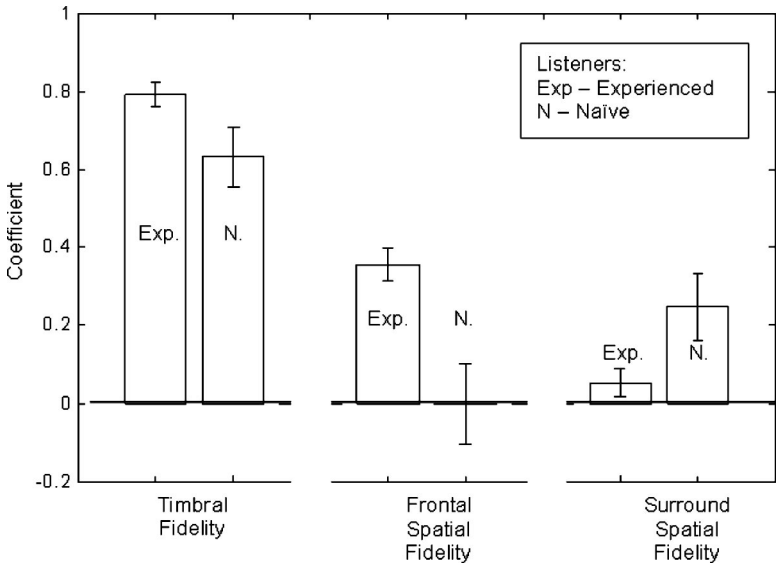


FIG. 5. Regression coefficients for the regression models predicting basic audio quality scores of experienced listeners and preferences of naïve listeners.

$$\text{BAQ} = 0.792\ \text{Timbral} + 0.356\ \text{Frontal} + 0.052\ \text{Surround} - 19.7, \quad (3)$$

where BAQ is the basic audio quality scores assessed by experienced listeners, Timbral is the timbral fidelity judged by experienced listeners, Frontal is the frontal spatial fidelity judged by experienced listeners, and Surround is the surround spatial fidelity judged by experienced listeners.

The second model (labeled as N in Fig. 5), described in detail in the preceding section, allows one to predict the preference scores of the naïve listeners using the scores of the timbral fidelity, frontal spatial, and surround spatial fidelities elicited from the experienced listeners. Its regression coefficients are given by Eq. (2).

The error bars depicted in Fig. 5 represent 95% confidence intervals. According to the plotted results, it is possible to note that both the scores from the experienced listeners and the preferences of the naïve listeners are primarily affected by the timbral fidelity (the largest coefficient). However, there is a substantial difference between the experienced and the naïve listeners in the weightings given to the frontal and surround spatial fidelity. The basic audio quality scores obtained from the experienced listeners (Exp) were influenced more by the frontal spatial fidelity (0.356) than by the surround spatial fidelity (0.052). On the contrary, the preferences of the naïve listeners (N) are only affected by the surround spatial fidelity (0.249). The contribution of the frontal spatial fidelity to the regression model in this case was proven to be statistically insignificant. Consequently, it can be concluded that the naïve listeners pay more attention to the surround channels and less attention' to the frontal spatial fidelity than experienced listeners.

IV. DISCUSSION

From the results presented it can be said that there was a high degree of similarity between expert and naïve ratings in general, but that naïve listeners tended to set more store by the presence of surround information than did experienced listeners when judging their preference. Furthermore, naïve listeners tended to assign no importance to frontal spatial fidelity in their preference ratings, whereas this was the spatial factor of greatest importance to the experienced listeners. This can almost certainly be attributed to the training that the experienced listeners had received in stereophonic listening, being strongly aware of issues relating to accurate phantom imaging in the front image, for example. Naïve listeners, on the other hand, are probably used to having one loudspeaker on the bookshelf and the other behind the couch (to use an anecdotal allusion), in which case their awareness of stereophonic phantom imaging would be likely to be negligible. Stereophonic reproduction, to many naïve listeners, may mean little more than a spatial effect, or high quality reproduction, in which case the presence or absence of generalized and immersive spatial effects would be more likely to be noticed by them and incorporated into their preferences. Furthermore, the novelty factor of surround sound is likely to be more prominent for naïve subjects than for experienced listeners.

V. CONCLUSIONS

At the outset the stated intention of this study was to discover what relationship might exist between the preferences of naïve listeners and the basic audio quality ratings of experienced listeners. This was investigated in relation to degraded multichannel audio items, limited in timbral and spatial quality by band limiting and down mixing. It was also hoped that some means of predicting consumer preference using expert judgements might be found. Finally the relationship between different fidelity attributes and consumer preference was sought.

In conclusion it can be said that there is a relatively large similarity between the basic audio quality scores acquired from the experienced listeners and preference scores elicited from the naïve listeners. The preference scores obtained from the naïve listeners can be, to some extent, approximated using solely the basic audio quality scores elicited from the experienced listeners, but the obtained regression model is

not very accurate (error equals 10%). Nevertheless, it can be useful for rough predictions of the preferences of naïve listeners based on the results of the basic audio quality scores acquired from experienced listeners. The preferences of naïve listeners could be predicted slightly more accurately by using three fidelity attributes graded by the experienced listeners as predictor variables (timbral fidelity, frontal spatial fidelity, and surround spatial fidelity). This suggests that expensive and time-consuming listening tests involving naïve listeners could possibly be replaced by less expensive tests incorporating experienced listeners, since the preferences of naïve listeners could be predicted, to a satisfactory degree of precision, using the scores obtained from experienced listeners. However it is not known whether this conclusion can be extended to include stimuli and degradation types not used here.

The fact that naïve listeners' preference can be predicted using the scores obtained from the experienced listeners does not imply that those two populations of listeners are similar in every aspect of audio quality evaluation. It was observed that the naïve listeners paid no attention to the frontal spatial fidelity, in contrast to the experienced listeners who considered this as important. It was also found that for the naïve listeners the surround spatial fidelity is a much more important factor than for the experienced listeners. However, the timbral fidelity dominates the preference judgements of both naïve and experienced subjects.

ACKNOWLEDGMENTS

This project was carried out with the financial support of the Engineering and Physical Sciences Research Council, UK. Some audio-visual excerpts used in the project were kindly supplied by BBC, R&D Department (used with permission).

Beerends, J. G., and De Caluwe, F. E. (**1999**). ''The influence of video quality on perceived audio quality and vice versa,'' J. Audio Eng. Soc. **47**, 335–362.

Diamantopoulos, A., and Schlegelmilch, B. B. (**1997**). *Taking the Fear out of the Data Analysis* (The Dryden Press, London).

Esbensen, K. (**2002**). *Multivariate Data Analysis—In Practice* (Camo Press, Oslo).

Field, A. (**2000**). *Discovering Statistics using SPSS for Windows* (Sage Publications, London).

Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Duxbury, New York).

ITU-R (**2001**). *Recommendation BS. 1534*: Method for the subjective assessment of intermediate quality level of coding systems. International Telecommunications Union.

Olive, S. (**2003**). ''Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: a case study,'' J. Audio Eng. Soc. **51**, 806–825.

Rumsey, F. (**2002**). ''Spatial quality evaluation for reproduced sound: terminology, meaning and a scene-based paradigm,'' J. Audio Eng. Soc. **50**, 651–666.

Zacharov, N., and Koivuniemi, K. (**2001a**). Audio descriptive analysis and mapping of spatial sound displays. Presented at 2001 International Conference on Auditory Display, Espoo, Finland, 29 July–1 August.

Zacharov, N., and Koivuniemi, K. (**2001b**). ''Unravelling the perception of spatial sound reproduction: Techniques and experimental design,'' in *Proceedings of the 19th AES International Conference. Surround Sound-Techniques, Technology and Perception. Schloss Elmau, Germany, 21–24 Jun* (Audio Engineering Society), pp. 272–286.

Zielinski, S., Rumsey, F., and Bech, S. (**2003a**). Comparison of quality degradation effects caused by limitation of bandwidth and by down-mix algorithms in consumer multichannel audio delivery systems. Presented at 114th AES Convention, Amsterdam, 22–25 March. Paper 5802. Audio Engineering Society.

Zielinski, S., Rumsey, F., and Bech, S. (**2003b**). ''Effects of bandwidth limitation on audio quality in consumer multichannel audio-visual delivery systems,'' J. Audio Eng. Soc. **51**, 475–501.

Zielinski, S., Rumsey, F., and Bech, S. (**2003c**). ''Effects of down-mix algorithms on quality of surround sound,'' J. Audio Eng. Soc. **51**, 780–798.

Zielinski, S. K., Rumsey, F., Kassier, R., and Bech, S. (**2005**). ''Comparison of basic audio quality, timbral and spatial fidelity changes caused by limitation of bandwidth and by down-mix algorithms in 5.1 surround audio systems,'' J. Audio Eng. Soc. **53**, 174–192.

Dead regions and pitch perception

Martina Huss and Brian C. J. Moore[a)]
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, United Kingdom

(Received 24 November 2004; revised 23 March 2005; accepted 29 March 2005)

The perception of pitch for pure tones with frequencies falling inside low- or high-frequency dead regions (DRs) was examined. Subjects adjusted a variable-frequency tone to match the pitch of a fixed tone. Matches within one ear were often erratic for tones falling in a DR, indicating unclear pitch percepts. Matches across ears of subjects with asymmetric hearing loss, and octave matches within ears, indicated that tones falling within a DR were perceived with an unclear pitch and/or a pitch different from ''normal'' whenever the tones fell more than 0.5 octave within a low- or high-frequency DR. One unilaterally impaired subject, with only a small surviving region between 3 and 4 kHz, matched a fixed 0.5-kHz tone in his impaired ear with, on average, a 3.75-kHz tone in his better ear. When asked to match the 0.5-kHz tone with an amplitude-modulated tone, he adjusted the carrier and modulation frequencies to about 3.8 and 0.5 kHz, respectively, suggesting that some temporal information was still available. Overall, the results indicate that the pitch of low-frequency tones is not conveyed solely by a temporal code. Possibly, there needs to be a correspondence between place and temporal information for a normal pitch to be perceived. [DOI: 10.1121/1.1920167]

PACS numbers: 43.66.Hg, 43.66.Sr [AJO] Pages: 3841–3852

I. INTRODUCTION

The pitch of a pure tone may be determined by the auditory system from the distribution of activity or excitation along the basilar membrane or in the auditory nerve (place theory) (Helmholtz, 1863; von Békésy, 1960; Siebert, 1968; 1970) or from temporal information derived from patterns of phase locking in the auditory nerve (Siebert, 1970; Moore, 1973; Goldstein and Srulovicz, 1977; Srulovicz and Goldstein, 1983). It remains unclear to what degree the different types of information are employed, what determines the pitch when either temporal or place information become weaker, and what happens when temporal and place codes give conflicting information about pitch.

In principle, theories of pitch perception can be evaluated by studying pitch perception for people with hearing impairment (Moore and Carlyon, 2005). In this paper, we examine the perception of pitch for tones whose frequencies fall inside dead regions (DRs) within the cochlea. A DR is a region where the inner hair cells (IHCs) and/or auditory neurons are functioning very poorly, or not at all. Moore (2004) proposed the following definition of a DR: ''a DR is a region in the cochlea where IHCs and/or neurons are functioning so poorly that a tone producing peak vibration in that region is detected by off-place listening (i.e., the tone is detected at a place where the amount of basilar-membrane vibration is lower, but the IHCs and neurons are functioning more effectively).'' The extent of a DR can be defined in terms of the CFs of the IHCs and/or neurons immediately adjacent to the DR (Moore, 2001). DRs can be diagnosed using psychophysical tuning curves (PTCs) (Florentine and Houtsma, 1983; Turner *et al.*, 1983; Moore and Alcántara, 2001; Huss and Moore, 2003; Kluk and Moore, 2005), or using the threshold-equalizing noise (TEN) test (Moore *et al.*, 2000; Huss and Moore, 2003; Moore *et al.*, 2004).

Studies investigating pitch perception in subjects with low-frequency DRs have suggested that a tone with a frequency falling in a DR is often perceived with a low pitch that is *roughly* ''normal'' (Florentine and Houtsma, 1983; Turner *et al.*, 1983). Pitch shifts were much smaller than would be predicted based on the place where the tone was assumed to be detected (just outside the boundary of the DR). The results were interpreted as indicating that the pitch of a low-frequency tone is predominantly derived from the temporal pattern of neural firing evoked by the tone. However, pitch shifts up to 30% were found in these studies. It remains unclear how these shifts are to be explained. The shifts may involve reduced precision of temporal information (Woolf *et al.*, 1981; Miller *et al.*, 1997), or a conflict between temporal and place information. Also, the variability of the matches was significantly greater for tones falling in the DR than for tones falling outside the DR, indicating that the pitch percept was less clear for the former. The subject of Florentine and Houtsma (1983), who had a low-frequency DR with an edge frequency of about 2 kHz, described tones with frequencies below 2 kHz as sounding ''hollow, without a body,'' ''very soft,'' and as if they were simultaneously combined with a ''narrow-band masker in the frequency range of the tone.''

The results obtained by Florentine and Houtsma (1983) and by Turner *et al.* (1983) were mostly based on pitch matches using tones at and within about 2 octaves of the edge frequency of the DR. It is possible that the analysis of information about pitch carried by interspike intervals is optimized when place and temporal information are consistent (Evans, 1978). For example, the analysis of temporal infor-

a)Author to whom all correspondence should be addressed.

mation may depend upon differences in the phase of the response at different points along the basilar membrane (Loeb *et al.*, 1983; Shamma and Klein, 2000; Carney *et al.*, 2002). The processing of temporal information could be disrupted when the propagation time of the traveling wave along the basilar membrane deviates from normal, as may happen in hearing-impaired ears (Ruggero *et al.*, 1996). When a tone produces peak vibration in a DR, perception of the tone depends on the spread of vibration to an adjacent functioning region of the cochlea. At that region, the traveling wave pattern, and specifically the relative phase of response at different places, might differ markedly from the pattern occurring around the place where peak vibration occurs. This might markedly disrupt the processing of temporal information.

The first aim of this paper was to investigate whether the ability of the auditory system to use temporal information for the perception of pitch weakens or disappears when the disparity between the frequency coded by the temporal pattern and the frequency corresponding to the peak of the excitation pattern increases. To do this, we determined the perceived pitch of tones with frequencies falling as much as 3 octaves inside a DR.

Previous studies of the perception of pitch by subjects with DRs have focused on subjects with low-frequency DRs. Here, we studied subjects with both low- and high-frequency DRs. Phase-locking information is markedly reduced for tones with frequencies above 5 kHz (Johnson, 1980; Palmer and Russell, 1986), although weak phase-locking information still may be available (Heinz *et al.*, 2001). It has not been demonstrated that high frequencies are coded solely by place information, although that is often assumed to be the case. If the pitch of tones with frequencies above 5 kHz is coded solely by place information, then the pitch of such tones should be determined by the place within the cochlea where they are transduced. For a tone falling in a DR, this place would correspond to one or other edge of the DR. A second aim of this study was to test whether the pitch of high-frequency tones falling in a DR was consistent with place coding.

II. EXPERIMENT 1: PURE-TONE MATCHING STIMULI

A. Procedure

1. Pitch-matching procedure

For the pitch-matching task, subjects were asked to match the perceived pitch of a variable pure tone with that of another fixed-frequency pure tone. The two tones were presented alternately. Matches were made across ears, to obtain a measure of diplacusis, and within one ear, to estimate the reliability of matching. However, within-ear matches were not obtained when subjects only had a limited time available for testing. Pitch shifts or pitch differences between ears (diplacusis) were estimated by measuring the frequency differences required to give ''equal pitch.'' The pitch matches were obtained using a method of adjustment. The listener used a three-button box. Button 3 was used to start and to finish the procedure for one match. Buttons 1 and 2 were used to adjust the frequency of the variable tone. The starting frequency was chosen randomly from the interval $0.5f$ to $1.5f$, where f denotes the frequency of the fixed tone. The subject controlled the direction of frequency change, using button 1 to increase the frequency and button 2 to decrease it. The frequency was altered between successive presentations of the variable tone, but not during presentation of the tone. If no button was pressed, the frequency was left unchanged. The term ''turnpoint'' is used to denote a point where the subjects reversed the direction of frequency change by changing from one button to the other. The frequency was initially changed in steps of 5% of the variable tone's frequency. After the fourth turnpoint, the step size was reduced to 1% to allow for a more precise final adjustment. The subject was instructed to ''bracket'' the point of equal perceived pitch, by making the variable tone sound alternately higher and lower in pitch than the fixed tone, before making the match. When the subject was satisfied with the match, this was indicated by pressing button 3. A matching run was aborted if the subject did not press button 3 by the 16th turnpoint. In an initial training session, at least one match was obtained for each fixed frequency. Matches for the training session were discarded. Subsequently, at least three matches were obtained for each fixed frequency. When time allowed, as many as eight matches were obtained.

2. Octave-matching procedure

For the octave-matching task, subjects were asked to adjust a tone of variable frequency so that it sounded 1 octave higher or lower than a fixed reference tone. The starting frequency was chosen randomly from the interval $1.5f$ to $2.5f$ (when asked to match an octave higher) or $0.25f$ to $0.75f$ (when asked to match an octave lower), where f denotes the frequency of the fixed tone. The subjects were sung and played the opening phrase of ''Somewhere over the rainbow,'' and told that the first two notes made an octave interval. The subjects were instructed to sing the octave aloud or subvocally to themselves during the matching procedure. From the group of ten subjects initially selected to do the task, only two subjects were able to perform the task reliably. In an initial training session, at least one match was obtained for each fixed frequency. Matches from the training session were discarded. At least three matches were obtained for each fixed frequency in subsequent sessions. When time allowed, as many as eight matches were obtained.

B. Stimulus generation

The stimuli were generated using a Tucker-Davis Technologies (TDT) System II, controlled by a computer. Stimuli were generated on-line for each trial and replayed through a 16-bit D/A converter (TDT DD1) at a sampling rate of 50 kHz. All stimuli were low-pass filtered at 20 kHz (Kemo VBF 8/04, 90-dB/oct slope). The levels of the fixed and variable-frequency tones were independently controlled by TDT PA4 programmable attenuators. For monaural presentation, the two tones were combined using a TDT SM3. Stimuli were then passed through a headphone buffer (TDT HB6) and one or two manual attenuators (Hatfield 2125) to the earphones. Subjects with reasonably symmetrical hearing

TABLE I. Audiometric thresholds (dB HL) and ages (years) for all subjects tested. Subject RC was completely deaf in her right ear. Asterisks indicate frequencies diagnosed as falling in a DR, using the TEN test and/or PTCs.

	Frequency (kHz)											
Subject (ear)	0.125	0.25	0.5	0.75	1	1.5	2	3	4	6	8	Age
AW(LE)	10	10	5	10	10	25	30	50	50*	55*	65*	84
AW(RE)	95*	85*	80*	75*	70*	75*	75*	60	80*	80*	90*	
CA(LE)	25	25	30	45	65*	115*						54
CA(RE)	20	15	35	55*	65*	115*						
DF(LE)	20	20	20	10	10	10	10		20	45	55	49
DF(RE)	75*	75*	80*	75*	70*	65*	60*	55*	60	90	90	
DV(LE)	10	15	10	5	5	5	5	35	55	60	65	41
DV(RE)	70	70	80*	75	75	75	70	75	75	80*	75	
MW(LE)	25	30	50	55	65*	85*	110*	110*				60
MW(RE)	20	25	55	60	70*	75*	100*	110*				
PJ(LE)	25	20	20	30	40	55	60	65	70*	75*	85*	69
PJ(RE)	15	10	10	30	35	60	60	70*	70*	75*	80*	
RC(LE)	15	10	10	5	5	60*	75*	90*	105*	115*		70

losses were tested using Sennheiser HD580 headphones. The maximum output level of the system, using the HD580 headphones, was about 112 dB SPL at 3 kHz. At this level, the harmonic distortion, measured using a KEMAR manikin (Burkhard and Sachs, 1975), was below 0.5%. Unilaterally impaired subjects, or subjects with highly asymmetrical losses, were tested with Etymotic Research ER2 insert earphones to give better attenuation across ears (Zwislocki, 1953). For these earphones, the harmonic distortion reached about 1% when the output level was 100 dB SPL. Subjects were tested in a double-walled sound-isolated booth.

All stimuli had a duration of 500 ms including 10-ms raised-cosine onset/offset ramps. The stimuli were presented in pairs, with the fixed-frequency tone followed by the variable-frequency tone. The interval between the stimuli within a pair was 500 ms, and the interval between pairs was 1000 ms.

To prevent subjects from basing their pitch matches on loudness cues, the tones were adjusted in level so as to give an approximately constant loudness level of either 50 or 60 phons. The lower loudness level was used when the sound-pressure levels needed to achieve a loudness level of 60 phons were higher than was possible with the equipment available at certain frequencies. Otherwise, the higher loudness level was used. The required adjustments in level were calculated using the loudness model described by Moore and Glasberg (1997). The model makes it possible to calculate loudness for tones of different frequencies and for different degrees of hearing loss. For the present application, the default parameter values of the model were used. The model takes into account the presence of any DR. The edge frequencies of the DRs specified for each subject were estimated from PTCs, as described by Huss and Moore (2003).

For each subject, 50- or 60-phon equal-loudness contours were calculated for frequencies from 20 to 12 000 Hz in 10-Hz steps. The model was used to calculate the SPL required at each frequency to give a fixed loudness level of 50 or 60 phons (corresponding to 2 or 4 sones). For subjects with a DR, the calculated equal-loudness contours did not always give equal perceived loudness when the tone frequency fell in a DR. For subjects AW and DV, the rate of growth of loudness with increasing sound level for tones within the DR, relative to the rate of growth for tones outside the DR, was greater than predicted by the loudness model. McDermott *et al.* (1998) reported loudness-matching data consistent with this finding. In these cases, levels were adjusted using feedback from the subject, for tones at the standard audiometric frequencies (including, where appropriate, 0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 6, and 8 kHz) until the loudness of the variable tone was consistently judged to equal that of the fixed tone. Interpolation (on a dB scale) was then used to determine the appropriate level of the variable tone at intermediate frequencies.

C. Subjects

1. Subjects for the pitch-matching task

The subjects were chosen from a larger group of subjects who had previously been tested using the TEN test and PTCs. For full details of how the dead regions were diagnosed, see Huss and Moore (2003). The edge frequencies, f_e, of the DRs were initially estimated using the TEN test, and for all subjects except one were defined more precisely from the tip frequencies of PTCs obtained when the signal frequency fell within a DR. In all cases where PTCs were obtained as well as TEN-test results, there was a good correspondence between the two; when the TEN-test criteria for a DR were met, the PTC showed a shifted tip. Most of the subjects tested here were also tested in an earlier study (Huss and Moore, 2003). The reader is referred to that paper to see examples of the correspondence between the TEN-test results and PTCs. Table I shows audiograms and estimated extents of DRs for all subjects. Three subjects (AW, DF, and DV) had asymmetric hearing loss, with low-frequency DRs of varying extent. AW and DV also had high-frequency DRs. One subject (RC) had one completely deaf ear and a high-frequency DR in the other ear. Three subjects (CA, MW, and PJ) had bilateral hearing losses with high-frequency DRs.

2. Subjects for the octave-matching task

Initially ten subjects were chosen from the larger group of subjects who had previously been tested using the TEN test and PTCs (Huss and Moore, 2003). Subjects were selected who thought of themselves as musical, or played an

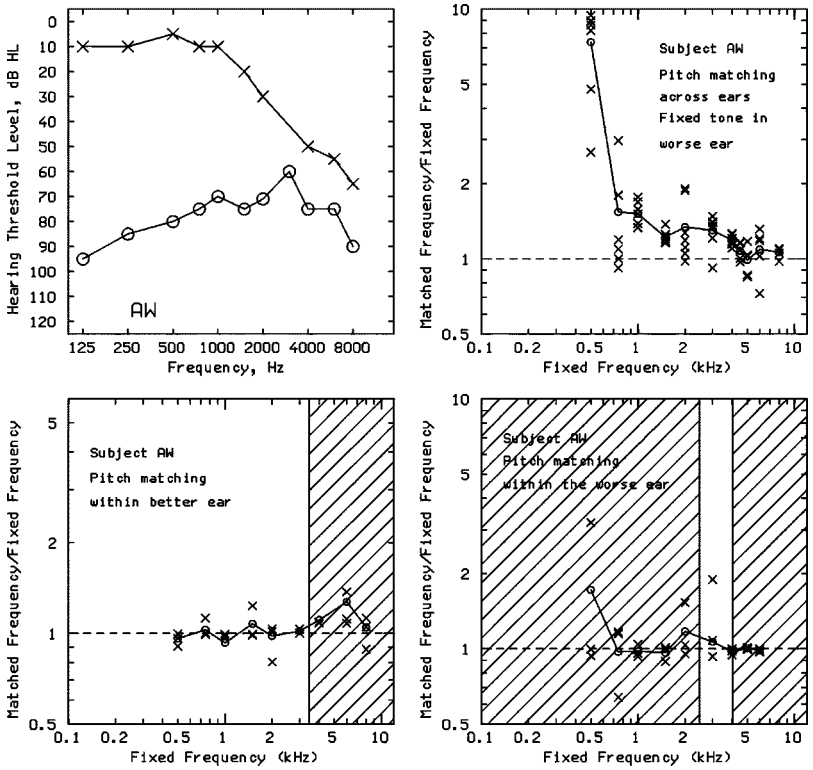


FIG. 1. Results for AW. His audiogram (top left) showed a moderate to severe hearing loss in his right ear, with minimum loss at about 3000 Hz (circles) and a high-frequency hearing loss in his left ear (crosses). AW had a low-frequency and high-frequency DR in his right ear, as indicated by the shaded areas in the bottom-right panel, and a surviving island centered at about 3500 Hz. The bottom panels show the results of pitch matching within the left ear (left panel) and right ear (right panel). The top-right panel displays results of pitch matching across ears. Each "×" denotes one match, and means are shown by open circles, connected by lines.

instrument, or sang. The subjects were instructed how to perform the octave-matching task and were given several hours of training using tones with frequencies where their hearing was comparatively good. Only two subjects from the group of ten subjects were able to perform the task reliably. One subject, DV, played the violin. He needed very little time to get accustomed to the task. The other subject, RC, needed about 8 h of training, and after a break of 2 months, another 3 h to relearn the task.

D. Results

Figures 1 to 7 display results for subjects AW, DF, DV, RC, CA, MW, and PJ, respectively. The ratio of the matched frequency to the fixed frequency is plotted as a function of the fixed frequency. Each×denotes one match, and means are shown by open circles connected by lines. The shaded area (if present) indicates a DR. Except for subject DF (Fig. 2), DRs are not shown for cases where matches were made across ears. Audiograms are also shown. Following audiological practice, absolute thresholds for the left ear are indicated by crosses and those for the right ear are indicated by circles. Pitch matches across ears were obtained for all hearing-impaired subjects except for RC, who is deaf in one ear. The scatter of the matches provides an indicator of frequency discrimination ability at each test frequency (Cardozo, 1965), which is probably related to the clarity of the perceived pitch.

1. Subjects with DRs that include low frequencies

We consider first the results for three subjects who had dead regions that included (but were not restricted to) low frequencies. Figure 1 shows results for subject AW. He had a moderate to severe hearing loss in his worse (right) ear (circles, top-left panel), with a surviving "island" around 3.5 kHz and extended DRs at lower and higher frequencies. In his left ear (crosses) he had a moderate high-frequency loss with a high-frequency DR starting at about 3.5 kHz. The pitch matches within his better ear (bottom left) were reasonably consistent at all frequencies. However, the mean match was somewhat shifted from a ratio of 1 for a fixed frequency of 6 kHz. It is surprising that AW was able to make reasonably accurate matches using his better ear even for high-frequency tones that fell within his DR. For frequencies above 5 kHz one would expect phase locking to be largely absent, and matching on the basis of place should not be accurate or consistent, as tones of all frequencies above f_e (3.5 kHz) should be detected at the same place. If the amount of excitation at that place was equated across frequency (because of the adjustments in level to equate loudness), it should not be possible to distinguish tones of different frequencies. Possibly, AW was able to use residual loudness cues to make matches at high frequencies. He was one of the subjects who reported that the levels selected using the loudness model did not lead to equal loudness of the fixed and variable tones. The subsequent adjustments made to the level of the variable tone may not have been sufficient to equate the loudness of the fixed and variable tones for all frequen-

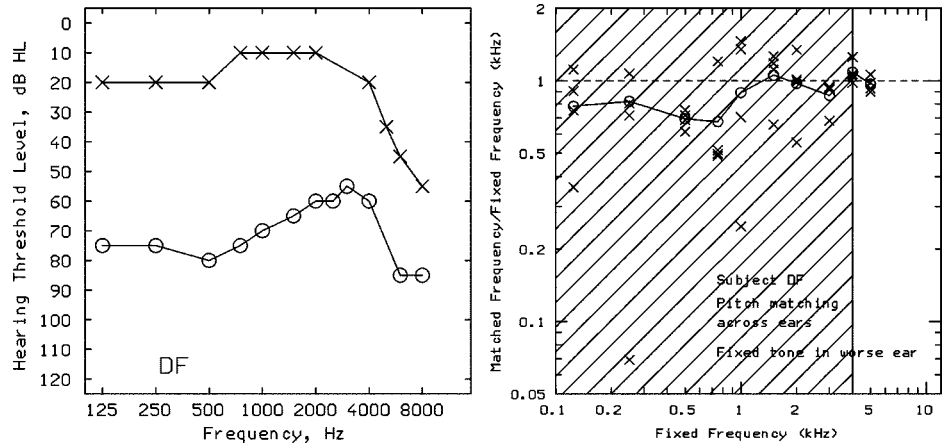


FIG. 2. Results for DF who has a low-frequency DR in her right ear (indicated by the shaded area in the right panel) and a high-frequency hearing loss in her left ear. The figure displays the results from pitch-matching across ears. Symbols have the same meaning as in Fig. 1.

cies, and AW may have adjusted the frequency of the latter so as to obtain a loudness match rather than a pitch match.

Matches within AW's worse ear (bottom right) were more erratic than for his better ear, indicating a less clear pitch percept. Matches were especially erratic for the lowest fixed frequency tested, 0.5 kHz, which fell well within his low-frequency DR. However, for 5- and 6-kHz tones, matches were surprisingly accurate, again perhaps reflecting the use of loudness cues.

Matches across ears, with the fixed tone in AW's worse ear (top right), showed considerable variability, but also some consistent deviations. A fixed 0.5-kHz tone in the worse ear was matched, on average, with a 3.75-kHz tone in the better ear (a ratio of about 7.5). Based on a *t*-test, the ratio of the matched frequency to the fixed frequency differed significantly from 1 ($p<0.01$). This result differs markedly from the results reported by Florentine and Houtsma (1983) and by Turner *et al.* (1983) for subjects with low-frequency DRs; their subjects seemed to hear roughly ''normal'' pitches for low-frequency tones falling within their DRs. Generally, for AW, the matched frequency lay above the fixed frequency, for all fixed frequencies up to about 4 kHz, indicating upward pitch shifts in the worse ear.

Figure 2 displays results for subject DF, who had one ear with a high-frequency hearing loss and one ear with a severe relatively flat hearing loss, with a low-frequency DR extending up to about 4 kHz. DF had limited time, and only pitch matches across ears were obtained, with the fixed tone in the worse ear. The matches showed large variability over the entire frequency range from 0.125 to 4 kHz, but became less erratic for the fixed frequency of 5 kHz, which fell above f_e. Matches were especially erratic when the fixed tone fell far inside the DR. A 1-kHz tone was matched with tones with frequencies ranging between 0.25 and 1.45 kHz, and a 0.25-kHz tone was matched with frequencies ranging between 0.025 and 0.275 kHz. The average matches fell within a range of ratios of 0.7 to 1.07. DF described tones with frequencies falling in the DR as sounding noise-like and difficult to compare with the clear percept of the pure tone presented to her left ear, but not particularly high pitched or ''screechy'' (a term used by one of the other subjects).

Results for subject DV are shown in Fig. 3. He has a high-frequency hearing loss in his left ear (crosses, top-left panel) and a moderate-to-severe flat loss in his right ear (circles), with localized DRs around 0.5 and 6 kHz. Because of time constraints, no within-ear matches were obtained for DV. Pitch matches across ears (top right), with the fixed tone in the worse (right) ear, showed good consistency and accuracy over the frequency range from 0.75 to 2 kHz. When the fixed tone fell in the DR around 0.5 kHz, matches were extremely erratic, with most matches falling above 0.5 kHz. However, the ratio of the matched frequency to the fixed frequency did not differ significantly from 1. When the fixed tone fell in or close to the DR around 6 kHz, matches were less erratic than for tones falling in the low-frequency DR. Matches at high frequencies deviated by up to 20% from the expected ratio. For example, a 5-kHz tone in the worse ear was matched with a 4-kHz tone in the better ear.

Octave matches within DV's better ear, with the upper tone fixed in frequency (bottom-left panel), showed ratios close to the ''expected'' value of 0.5 for frequencies of the fixed tone up to 4 kHz. Matches became more erratic, with some systematic shifts, when the frequency of the fixed tone was above 5 kHz. Similar but less marked effects have been reported for normally hearing subjects, and attributed to a loss of the sense of ''tone chroma'' at high frequencies (Ward, 1954).

Octave matches within DV's worse ear, with the upper tone fixed in frequency (bottom-right panel), showed much more variability, indicating unclear pitch perception, when either the fixed or the variable tone fell in one of the DRs. Many matches were made with ratios markedly shifted from 0.5. For example, the fixed 0.5-kHz tone was matched with a 0.45-kHz tone instead of with a 0.25-kHz tone. The ratio of the matched frequency to the fixed frequency differed significantly from 0.5 ($p<0.01$). Similarly, the fixed 0.44-kHz tone was matched with a 0.3-kHz tone instead of with a 0.22-kHz tone. The ratio of the matched frequency to the fixed frequency differed significantly from 0.5 ($p<0.01$). Only the fixed 0.22-kHz tone and the fixed 2-kHz tone were matched with a variable tone close to a ratio of 0.5. These were the only cases where both the fixed and the variable tones had frequencies falling outside either DR. The erratic and shifted matches for the other cases indicate that tones

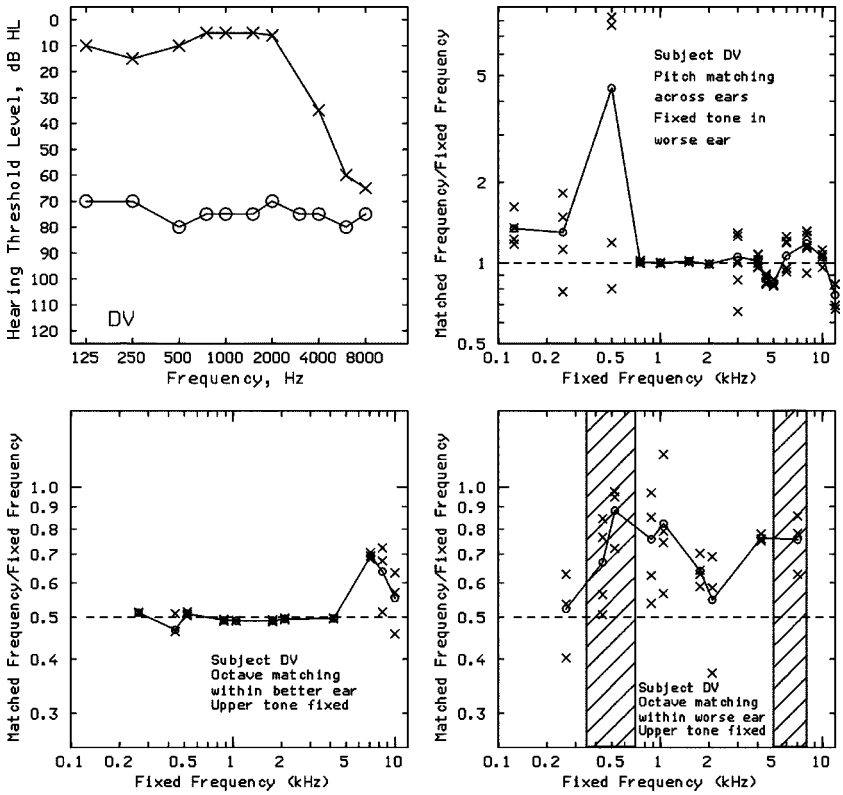


FIG. 3. Results for DV, whose audiogram (top left) shows a relatively flat hearing loss in his right ear (circles) and a high-frequency hearing loss in his left ear (crosses). DV has DRs around 0.5 and 6 kHz in his right ear, as indicated by the shaded areas in the bottom-right panel. The top-right panel displays the results of pitch matching across ears. The two bottom panels display the results of the octave-matching task. Shading and symbols have the same meaning as in Fig. 1.

falling within one of the DRs were not perceived with normal pitches.

In summary, tones falling within a low-frequency DR were sometimes perceived with a nearly normal pitch, but were often perceived with an upward-shifted pitch, especially when the tones fell well within the DR. Matches were often erratic for tones falling within a low-frequency DR, indicating unclear pitch sensations.

2. Subjects with DRs restricted to medium and high frequencies

We consider now the results for subjects whose DRs were restricted to medium and high frequencies. Results for RC are shown in Fig. 4. RC was diagnosed with a high-frequency DR with f_e at about 1.1 kHz, as indicated by the shaded areas (the other ear was deaf). The top-right panel shows the results of the within-ear pitch-matching task. The matches were relatively accurate for frequencies between 0.25 and 1.25 kHz, but became more erratic for tones with frequencies above 1.75 kHz, indicating that a clear pitch percept was not obtained at frequencies above 1.75 kHz. However, the mean matched frequency differed by less than 10% from the frequency of the fixed tone at 1.75 and 2 kHz. Even though RC matched tones more erratically within the DR, her average pitch match was nevertheless reasonably close to a ratio of 1. It seems likely that all tones with frequencies above 1.2 kHz were detected at roughly the same place in the cochlea, i.e., the place with CF just below 1.1 kHz. If pitch were determined solely by place, then one would expect all tones with frequencies above 1.1 kHz to have the same pitch. The fact that the mean matches were reasonably accurate for frequencies up to 2 kHz suggests that pitch was not determined solely by place information.

The two bottom panels show octave-matching results. When the lower tone was fixed in frequency (bottom right), then for frequencies up to 0.75 kHz the frequency ratio of the variable and the fixed tones was reasonably close to 2:1. For a fixed frequency of 1 kHz, the upper tone was adjusted to about 1.4 kHz, corresponding to an interval much less than an octave. The ratio for this fixed frequency differed significantly from 2 ($p<0.01$). When the upper tone was fixed in frequency (bottom left), the frequency ratios were around, but a little above, 0.5, the ''expected'' value, for fixed frequencies up to 1 kHz. For fixed frequencies of 1.75 and 2 kHz, which fell within the DR, the ratios were markedly greater than 0.5. The ratios for these two fixed frequencies deviated significantly from 0.5 ($p<0.01$ in both cases). Thus, the resulting interval was less than an octave. Despite the high variability of the pitch matches, tones with frequencies falling in the DR were thus consistently matched as if they were perceived with an abnormally high pitch.

It proved impossible to obtain pitch or octave matches for RC when the frequency of one of the tones fell above 2 kHz. The subject's responses became very erratic and sometimes she attempted to adjust the variable frequency to very high values, where her absolute thresholds increased. The higher frequencies were accordingly associated with high levels, which caused the program to interrupt the procedure.

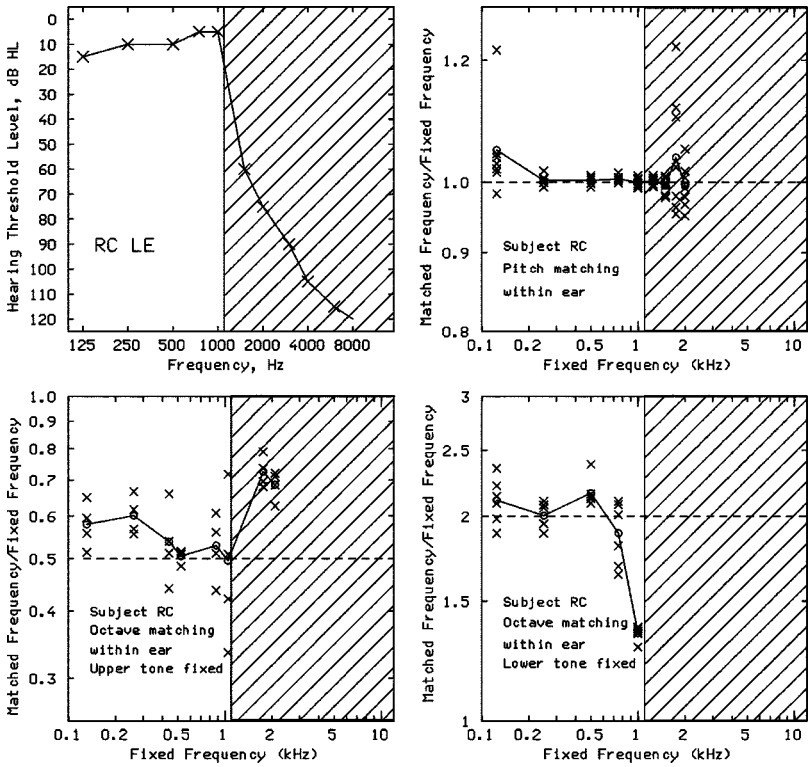


FIG. 4. Results for RC, who has a high-frequency DR (f_e at about 1.1 kHz), as indicated by the shaded area. RC has a steeply sloping audiogram (indicated by crosses connected by lines in the top-left panel). RC is deaf in her right ear. The top-right panel shows the results for the pitch-matching task. The two bottom panels display the results from the octave-matching task. Symbols have the same meaning as in Fig. 1.

Figure 5 displays results for subject CA. The subject has a bilateral steeply sloping audiogram (top left) and a high-frequency DR in both ears with f_e at 0.85 kHz for the left ear and at 0.5 kHz for the right ear. The pitch matches obtained within ears (bottom panels) are relatively precise for the lower frequencies and become less precise for tones with frequencies above 0.75 kHz in both ears. Despite CA's ability to match tones within ears with relatively little variability, his pitch matches across ears (top right) were very erratic (about 20% above and below the mean value) over the entire frequency range tested. Although the audiograms for the two ears are similar, the frequency-to-place map might differ across the two ears. This would make it impossible to match both place and temporal information across the two ears. It is impossible to tell from these data whether CA perceives pitches that are different from normal. He was unable to perform octave matching, and the lack of systematic deviations across ears (top-right panel) might reflect similar pitch shifts in the two ears.

Figure 6 displays results for subject MW, who had a bilateral hearing loss and a high-frequency DR in each ear, with f_e about 0.75 kHz. Because of time constraints, no within-ear matches were obtained for MW. For pitch matches across ears, the average frequency ratio was close to 1 for low frequencies, but fell increasingly below 1 once the frequency of the fixed tone was above 0.3 kHz. Even tones with frequencies falling outside the DR were matched with large variability, but the mean ratio for tones falling outside the DR did not fall below 0.9. The ratio decreased to about 0.75 for a 0.9-kHz tone, which fell within the DR. For the fixed frequency of 0.9 kHz, the ratio was significantly below 1 ($p<0.01$), indicating a shift of pitch in one ear relative to the other, despite the symmetrical nature of the hearing loss. It was not possible to obtain matches for higher frequencies, owing to the very high sound levels that would have been required.

Figure 7 shows results for subject PJ, who also had a bilateral hearing loss and a high-frequency DR in each ear with f_e about 3 kHz in the right ear (bottom left) and 3.5 kHz in the left ear (top left). Within-ear matches were obtained only for the right ear. Results for that ear (bottom right) showed very good consistency for frequencies of 0.25 and 0.5 kHz, but matches became somewhat more erratic with increasing frequency, and became very erratic when the frequency of the fixed tone fell within the DR. It was not possible to obtain matches for fixed frequencies above 4 kHz. The pitch matches across ears for subject PJ resulted in ratios relatively close to 1, indicating no large systematic shifts in pitch in one ear relative to the other. However, the matches became very erratic when the fixed frequency was 3 kHz. It was impossible to obtain matches for frequencies above 3 kHz, owing to the very high sound levels that would have been required.

In summary, pitch matches were often very erratic for tones falling in a high-frequency DR, indicating unclear pitch percepts. Pitch shifts were not always observed, but one subject (LE) showed substantial shifts even when the frequency of the tone fell only a little inside the DR, and one subject (MW) showed differences in pitch across ears despite a symmetrical hearing loss.

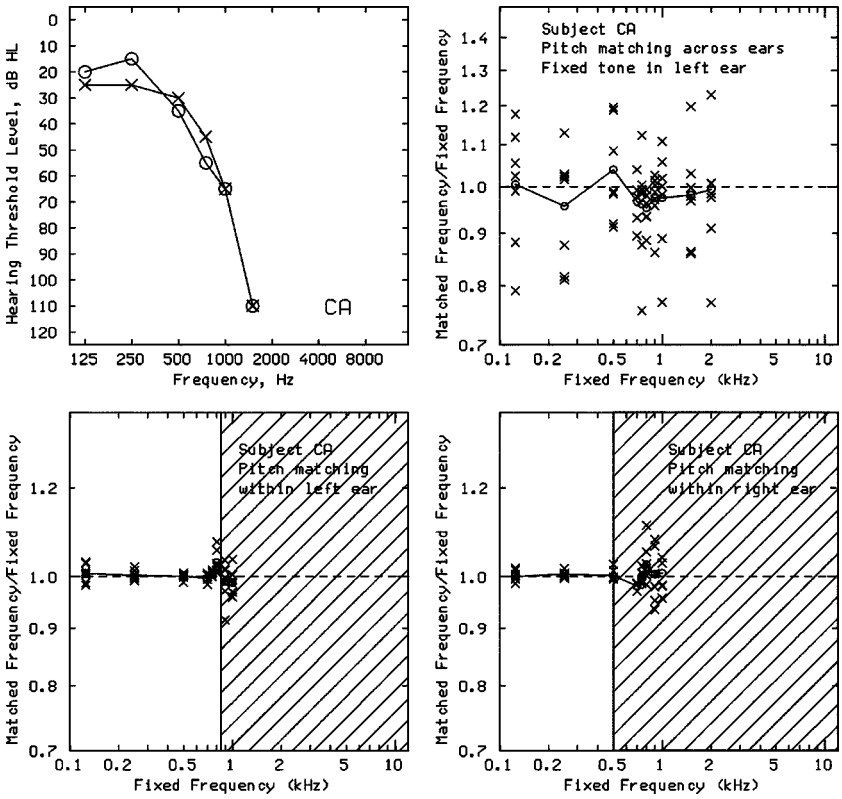


FIG. 5. Results for CA, who has high-frequency DRs in his left and his right ears, as indicated by the shaded areas in the bottom panels. The bottom panels display results of pitch matching within ears. The top-right panel displays the results of pitch matching across ears. Shading and symbols have the same meaning as in Fig. 1.

III. EXPERIMENT 2: AMPLITUDE-MODULATED TONES AS MATCHING STIMULI

A. Rationale

The pitch matches across ears for subject AW, with the fixed tone in his worse ear, showed considerable variability, but also some large consistent deviations for tones with frequencies falling about 3 octaves inside his low-frequency DR. A fixed 0.5-kHz tone in his worse ear was matched, on average, with a 3.75-kHz tone in his better ear. This is consistent with the ''place'' of the surviving functioning ''island'' (centered at about 3.5–4 kHz) in his worse ear. AW complained that it was difficult to match the 0.5-kHz tone, presented to his worse ear, which he perceived as a ''screeching high-pitched noise.'' There seemed to be no pure-tone frequency in his better ear that gave a good match. The great difficulty in matching the pitch of tones with very different sound quality is well documented (in terms of large variability) for tones with very different spectra (Plomp, 1976; Moore and Glasberg, 1990). For AW, the difficulty may have occurred because it was not possible to match both place and temporal information across ears. If the 0.5-kHz tone in the worse ear was transduced by IHCs with CFs of about 4 kHz, then discrepancies in place across ears would be minimized when the matching frequency in the better ear was about 4

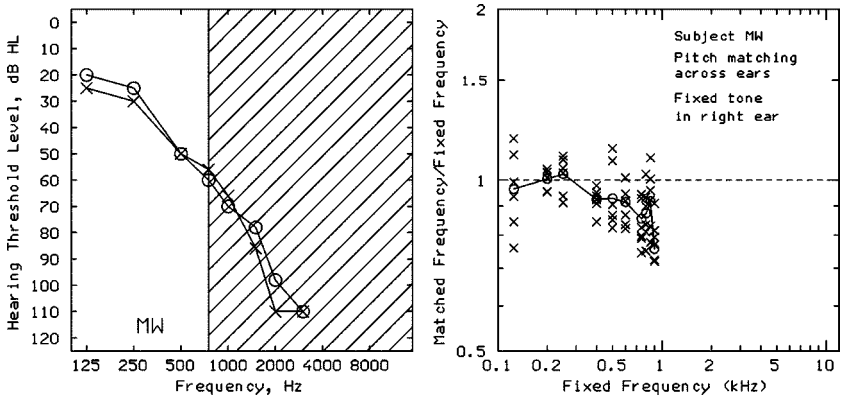


FIG. 6. Results for MW, who had DRs in both ears starting at 750 Hz. The right panel shows results of pitch matching across ears. Symbols have the same meaning as in Fig. 1.

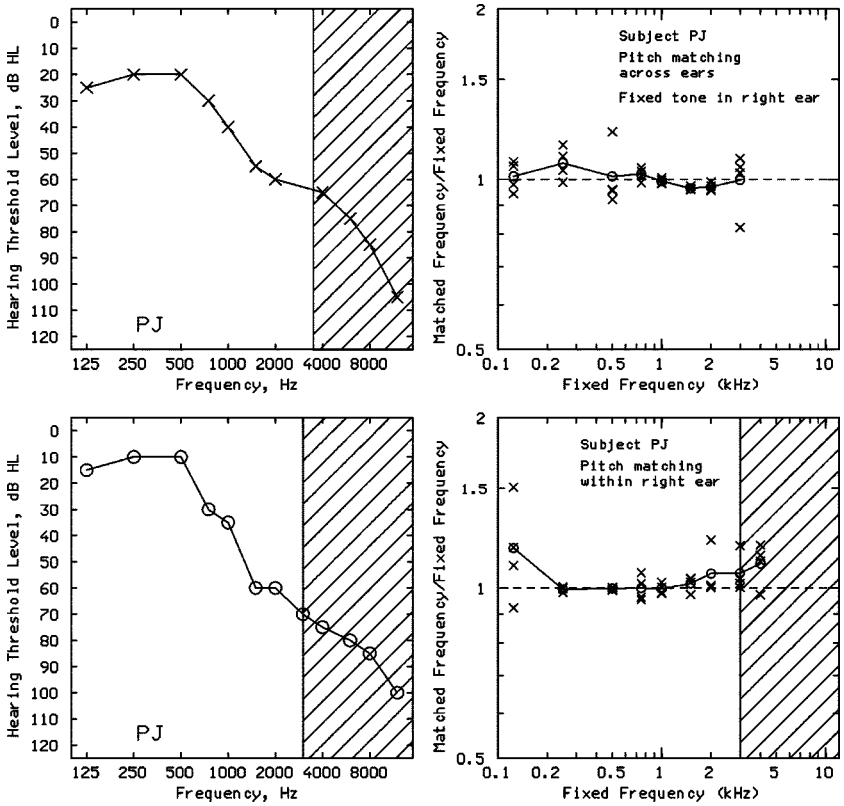


FIG. 7. Results for PJ, who has high-frequency DRs starting at 3500 Hz in the left ear (top-left panel and at 3000 Hz in the right ear (bottom left). The bottom-right panel shows results of pitch matching within the right ear, and the top-right panel shows results of pitch matching across ears. Symbols have the same meaning as in Fig. 1.

kHz. However, the temporal information would then differ considerably across the two ears.

Since the sensation evoked by a tone falling in a DR may depend on both place and temporal information, it was decided to explore whether a better match was possible when AW could vary place and temporal information independently from one another. This was achieved through the use of a matching stimulus consisting of a sinusoidally amplitude-modulated (SAM) sinusoidal carrier. The carrier frequency determined the place of excitation and the modulation frequency determined the temporal pattern at that place (Oxenham *et al.*, 2004). SAM tones can evoke both a ''place pitch'' or ''spectral pitch'' corresponding to the carrier frequency and a lower pitch, sometimes called residue pitch (Schouten, 1940; Ritsma, 1962; 1963), which is related to the modulation frequency. The perception of the low pitch is probably derived from temporal information in the envelope of the waveform, provided the ratio of carrier to modulation frequency is high (Moore and Moore, 2003a; 2003b). However, the envelope-based pitch produced by amplitude modulation is less salient than the pitch elicited by pure tones (Ritsma, 1962; 1963; Burns and Viemeister, 1976; 1981; Moore and Moore, 2003a; 2003b).

In the experiment, AW was allowed to adjust both the carrier frequency and the amplitude modulation frequency of the variable stimulus to give as close a match as possible to what he perceived in his worse ear.

B. Method

A fixed 0.5-kHz pure tone was presented to the worse ear and an SAM tone with modulation depth of 50% was presented to the better ear. The waveform, $T(t)$, of the SAM tone was defined by

$$T(t)=[1+0.5\sin(2\pi f_m t)]\sin(2\pi f_c t), \quad (1)$$

where f_m is the modulation frequency and f_c is the carrier frequency. The timing of the stimuli was the same as used in the pitch-matching task. The subject was instructed to adjust f_m and f_c in turn, using buttons on the response box, so as to achieve a match. The subject was able to use a third button to switch between controlling f_m and controlling f_c. A fourth button was used to select either large step sizes (20% for the modulation frequency and 5% for the carrier frequency) or small step sizes (1% for the modulation frequency and 0.5% for the carrier frequency). AW was allowed to switch as often as required to be satisfied with the result. Six matches were obtained. Stimuli were generated in the same way as before, and were delivered using ER2 insert earphones.

TABLE II. Means and standard deviations of six matches between a 0.5-kHz pure tone in AW's worse ear and an adjustable SAM tone in his better ear.

	Matched carrier frequency (kHz)	Matched modulation frequency (kHz)
Mean	3.8	0.52
SD	1.2	0.10

C. Results

Although AW spent a considerable time exploring various combinations of modulation and carrier frequency, so that each match took a long time, he stated that he was more satisfied with the final matches obtained than he was when he could only adjust the frequency of a pure tone, as in the first experiment. The results obtained using the SAM matching stimulus are summarized in Table II. AW matched the fixed 0.5-kHz pure tone with a SAM tone with $f_c \approx 3.8$ kHz and $f_m \approx 0.52$ kHz. The standard deviation of the matches was relatively large, especially for f_c.

The results clearly indicate that some temporal information about the fixed tone in the worse ear was preserved in the auditory system and influenced the percept. The results are consistent with the idea that the 0.5-kHz fixed tone in AW's worse ear evoked a ''place pitch'' corresponding to the CF of the small surviving region in his worse ear, and a ''temporal pitch'' related to the temporal fine structure of the fixed tone. This temporal pitch could be matched by adjusting the modulation frequency of the SAM tone.

IV. DISCUSSION

The results show that a tone with frequency falling in a DR is often but not always perceived with a different pitch from normal. The deviations, expressed as the ratio of the frequency of the matching tone to the ''expected'' value, were often greater than 1.2 and ranged up to a factor of 7. The shifts tended to be especially large when the frequency of the fixed tone fell well within the DR. However, subject DV showed a substantial shift when the frequency of the fixed tone fell within a restricted low-frequency DR, and RC showed a substantial shift when the frequency fell just inside a high-frequency DR. In addition, pitch matches were often very erratic for tones with frequencies falling in a DR, indicating that the tones did not evoke clear pitches.

Previous studies of the pitch of tones falling in DRs (Florentine and Houtsma, 1983; Turner *et al.*, 1983) have been interpreted as indicating only small shifts in pitch. However, we believe that there are not in fact any large discrepancies across studies. The single unilaterally impaired subject tested by Florentine and Houtsma consistently showed upward pitch shifts of the tone in the low-frequency DR when matching across ears, although the pitch clearly did not correspond to the place where the tone was detected. Results comparable to this were found for one of our subjects, DF, except that DF tended to show slight downward shifts in pitch and the matches showed greater variability for DF than for the subject of Florentine and Houtsma. One unilaterally hearing-impaired subject tested by Turner *et al.* had only a very restricted low-frequency DR, with f_e at about 0.5 kHz. That subject did not show clear pitch shifts when matching across ears, but the fixed tone in the impaired ear never fell more than 1 octave below f_e. Another subject tested by Turner *et al.* had bilateral hearing loss with a low-frequency DR in the test ear with $f_e \approx 0.8-1$ kHz. Octave-matching results did not show clear pitch shifts (shifts in the mean matches ranged from 5% to -2% relative to the expected values). However, the frequencies were chosen so that both of the tones (i.e., the fixed tone and the variable tone) fell within the DR or close to the boundary of the DR. Therefore, both tones could have been shifted in pitch from normal values by roughly the same factor, leading to a ''preserved'' octave relationship.

For our data, the deviations in the pitch matches in most cases (RC, AW, DV), but not in all cases (DF), indicated that the perceived pitch was shifted upward, regardless of whether the tone fell in a low-frequency or a high-frequency DR. In the case of AW, a 0.5-kHz tone in his worse ear (well within the DR) was matched, on average, with a 3.75-kHz tone in his better ear. This may indicate that the pitch percept was dominated by the place of transduction, because AW has only a surviving island of IHCs between 3 and 4 kHz in his worse ear. AW described the 0.5-kHz tone as sounding screechy and very high pitched. However, the results for RC are not easy to explain in terms of the place of transduction. RC has a high-frequency DR starting at about 1.1 kHz. For the octave-matching task, when the upper tone fell within her DR, the matches indicated that she perceived the tone falling in the DR with an abnormally high pitch. If the pitch of that tone was influenced by the place of transduction, then the pitch would be expected to be lower, not higher, than normal.

For tones with frequencies above 5 kHz, phase locking is very weak (Johnson, 1980; Palmer and Russell, 1986). Therefore, one would expect that, for tones whose frequencies fall in a high-frequency DR with f_e above 5 kHz, the place of transduction would be the major factor influencing pitch. The results for DV could be interpreted in this way. He had a restricted high-frequency DR centered around 6 kHz. A tone with frequency falling toward the upper edge of the DR was matched with a tone in the other ear at a higher frequency, consistent with it being transduced at the functioning place in the cochlea above the DR. Conversely, a tone with frequency falling toward the lower edge of the DR, was matched with a tone in the other ear at a lower frequency, consistent with it being transduced at the functioning place in the cochlea below the DR. The 6-kHz tone, which fell in the middle of the relatively small DR, might have been transduced by IHCs both above and below the DR, and this could account for the very small shift in pitch observed in this case. Thus, the results are broadly consistent with the idea that place information determined the pitch shifts.

Unfortunately, it was not possible to obtain clear-cut results using tones with frequencies above 5 kHz for other subjects with high-frequency DRs,. The subjects with high-frequency DRs who were able to hear such tones (PJ, MW, and AW, right ear) all had bilateral DRs at high frequencies, so no estimate of the perceived pitch could be obtained by comparison with a normal ear. Also, these subjects did not learn to match octaves reliably.

In the Introduction, we discussed the possibility that there needs to be a correspondence between place and temporal information in order for a normal pitch to be perceived. It is possible that temporal coding is sufficient to convey a near-normal pitch when the discrepancy between place and temporal information is small, but that the coding of pitch via temporal information is disrupted when the discrepancy is large, as would occur when the frequency of a tone falls

well inside a DR. This idea can be tested by considering the data for AW. In his case, matched tone frequencies clearly diverged increasingly from the frequency of the fixed tone in his worse ear with increasing distance of the tone frequency from the boundary, f_e, of the DR. However, shifts of more than 30% occurred even when the frequency of the fixed tone fell only just below f_e. A 0.5-kHz tone in his worse ear, which fell more than 2.5 octaves below f_e, was matched with a 3.75-kHz tone in his better ear, which corresponds roughly to f_e. Thus, when the fixed tone fell well within the DR, it seems as if the pitch of the tone was matched according to the place of transduction, independently of temporal information.

As noted earlier, AW reported that it was very difficult to match a fixed 0.5-kHz tone in his worse ear with a pure tone in his better ear. Even when he had made a match, he reported that the sounds in the two ears were perceived differently. The results of the task using an SAM matching stimulus showed that some temporal information was preserved and that it did influence the percept. Even though AW stated that we was more satisfied with the matches using the SAM tone than using the pure tone, the matches with the SAM tone were still somewhat erratic. This may reflect the fact that, even for normally hearing subjects, the ''low'' pitch of a 4-kHz carrier, with 50% modulation depth and 0.5-kHz modulation frequency is comparatively weak (Ritsma, 1962; 1963). Nevertheless, the fact that he was able to adjust the modulation frequency of the SAM tone so as to approximately match the frequency of the fixed sinusoid in his worse ear clearly indicates that temporal information was conveyed to some extent, even though the IHCs transducing the 0.5-kHz tone had CFs more than 2.5 octaves higher than would normally be the case.

Another subject (DF) with a unilateral low-frequency DR did not show large pitch shifts when matching across ears, even when the frequency of the fixed tone fell well inside the DR. The pitch matches when the fixed tone fell within the DR were very erratic, but they did not show a progressively increasing pitch shift with increasing distance from the boundary of the DR. For example, a 0.25-kHz tone, which fell 4 octaves below f_e, was matched with a frequency of about 0.20 kHz, a ratio of 0.8. Also, the variability of the matches did not increase systematically as the frequency of the fixed tone was moved further below f_e. The growing mismatch between place and temporal information did not seem to lead to an increasing degradation of the pitch percept for DF. The results for DF suggest that a mismatch between place and temporal information does not always lead to a large pitch shift, although it does lead to a much less clear pitch percept.

As described earlier, a sinusoidal tone with frequency below 5 kHz may evoke both a place pitch related to the place where the tone is detected, and a temporal pitch related to the pattern of phase locking evoked by the tone. These two types of pitch percept have been proposed to account for results obtained from people with cochlear implants (McKay *et al.*, 1994; 2000; Moore and Carlyon, 2005). The relative salience of these two pitches may vary across hearing-impaired individuals. For AW, the place pitch may have been the dominant percept, since his pitch matches for the sinusoidal fixed stimuli seemed to have been based largely on the place of transduction of the fixed tone. For DF, the temporal pitch may have been more dominant. The relative salience of the temporal pitch in a specific subject may depend partly on the status of the IHCs that are responsible for transduction of the tone. If the IHCs are functioning poorly, the temporal pattern of neural spikes evoked by the tone may be relatively sparse, leading to a poor temporal representation, and a weak temporal pitch. This idea is comparable to our earlier proposal that tone decay (adaptation) is related to the status of the IHCs responsible for transduction of the tone (Huss and Moore, 2003). In this context, it is noteworthy that the subject tested by Florentine and Houtsma (1983), who produced reasonably accurate pitch matches for tones falling within the DR, had an absolute threshold at f_e of about 40 dB SPL (corresponding to about 30 dB HL), which probably indicates reasonably good IHC function just outside the DR. That may account for the apparent dominance of temporal pitch for the subject of Florentine and Houtsma.

In our data, the pitch of tones falling in a DR was often higher than normal, regardless of the location of the DR (high- or low-frequency DR). This effect could be a result of the reduced precision of phase locking produced by damaged IHCs at the boundary of the DR. A tone that would normally produce strong phase locking might produce little or no phase locking when transduced by damaged IHCs. The resulting temporal pattern could resemble that elicited by a higher frequency tone, because higher frequencies are associated with less precise phase locking (Johnson, 1980; Palmer and Russell, 1986). Both place and temporal information would still be perceived to some degree (as in the case for AW), but there may also be a tendency for the more erratic temporal pattern to lead to a perceived pitch which is somewhat higher than normal.

V. CONCLUSIONS

(1) Pitch matches within one ear were often erratic for tones falling more than half an octave into a low-frequency or high-frequency DR. This indicates that such tones usually do not evoke a clear pitch sensation.
(2) Pitch matches across the ears of subjects with asymmetric hearing loss and octave matches within ears indicate that tones falling within a DR are often perceived with a pitch different (usually higher) than normal. This was true for tones falling in both low-frequency and high-frequency DRs.
(3) The results indicate that the pitch of low-frequency tones is not represented solely by a temporal code. Possibly, there needs to be a correspondence between place and temporal information for a normal pitch to be perceived.
(4) One subject with highly asymmetric hearing loss made matches of a 0.5-kHz pure tone falling in a DR in the worse ear with an SAM tone in the other (better) ear. The carrier frequency of the SAM tone was adjusted to about 3.8 kHz and the modulation frequency was adjusted to about 0.52 kHz. The matched carrier frequency corresponded roughly with the place where the tone in the worse ear was detected, while the matched modulation

frequency corresponded roughly with the frequency of the tone in the worse ear, suggesting that some temporal information was preserved.

(5) The shifts in the matches for tones above 5 kHz, when those frequencies fell in a high-frequency DR, were roughly consistent in extent and direction with what would be expected if the pitches were solely coded by the place of transduction.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK), DAAD (Germany), and BBSRC (UK). We thank Brian Glasberg, Tom Baer, Josephine Marriage, Debi Vickers, and Michael Stone for their contributions to various aspects of this work. We thank Chris Darwin for suggesting the use of an SAM matching stimulus. We thank Andrew Oxenham and two anonymous reviewers for their comments on an earlier version of this paper.

Burkhard, M. D., and Sachs, R. M. (**1975**). ''Anthropometric manikin for acoustic research,'' J. Acoust. Soc. Am. **58**, 214–222.

Burns, E. M., and Viemeister, N. F. (**1976**). ''Nonspectral pitch,'' J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). ''Played again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Cardozo, B. L. (**1965**). ''Adjusting the method of adjustment. SD vs DL,'' J. Acoust. Soc. Am. **37**, 786–792.

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (**2002**). ''Auditory phase opponency: A temporal model for masked detection at low frequencies,'' Acust. Acta Acust. **88**, 334–346.

Evans, E. F. (**1978**). ''Place and time coding of frequency in the peripheral auditory system: Some physiological pros and cons,'' Audiology **17**, 369–420.

Florentine, M., and Houtsma, A. J. M. (**1983**). ''Tuning curves and pitch matches in a listener with a unilateral, low-frequency hearing loss,'' J. Acoust. Soc. Am. **73**, 961–965.

Goldstein, J. L., and Srulovicz, P. (**1977**). ''Auditory-nerve spike intervals as an adequate basis for aural frequency measurement,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). ''Evaluating auditory performance limits. I. One-parameter discrimination using a computational model for the auditory nerve,'' Neural Comput. **13**, 2273–2316.

Helmholtz, H. L. F. (**1863**). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg, Braunschweig).

Huss, M., and Moore, B. C. J. (**2003**). ''Tone decay for hearing-impaired listeners with and without dead regions in the cochlea,'' J. Acoust. Soc. Am. **114**, 3283–3294.

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

Kluk, K., and Moore, B. C. J. (**2005**). ''Factors affecting psychophysical tuning curves for hearing-impaired subjects,'' Hear. Res. **200**, 115–131.

Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). ''Spatial cross correlation: A proposed mechanism for acoustic pitch perception,'' Biol. Cybern. **47**, 149–163.

McDermott, H. J., Lech, M., Kornblum, M. S., and Irvine, D. R. F. (**1998**). ''Loudness perception and frequency discrimination in subjects with steeply sloping hearing loss; Possible correlates of neural plasticity,'' J. Acoust. Soc. Am. **104**, 2314–2325.

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.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). ''Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees,'' J. Acoust. Soc. Am. **96**, 2664–2673.

Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). ''Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers,'' J. Acoust. Soc. Am. **101**, 3602–3616.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

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. (**2004**). ''Dead regions in the cochlea: Conceptual foundations, diagnosis, and clinical applications,'' Ear Hear. **25**, 98–116.

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 Carlyon, R. P. (**2005**). ''Perception of pitch by people with cochlear hearing loss and by cochlear implant users,'' in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York).

Moore, B. C. J., and Glasberg, B. R. (**1990**). ''Frequency discrimination of complex tones with overlapping and nonoverlapping harmonics,'' J. Acoust. Soc. Am. **87**, 2163–2177.

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 Moore, G. A. (**2003a**). ''Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects,'' Hear. Res. **182**, 153–163.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**2004**). ''New version of the TEN test with calibrations in dB HL,'' Ear Hear. **25**, 478–487.

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, G. A., and Moore, B. C. J. (**2003b**). ''Perception of the low pitch of frequency-shifted complexes by normal-hearing listeners,'' J. Acoust. Soc. Am. **113**, 977–985.

Oxenham, A. J., Bernstein, J. G., and Penagos, H. (**2004**). ''Correct tonotopic representation is necessary for complex pitch perception,'' Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guineapig and its relation to the receptor potential of inner hair-cells,'' Hear. Res. **24**, 1–15.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

Ritsma, R. J. (**1962**). ''Existence region of the tonal residue. I.,'' J. Acoust. Soc. Am. **34**, 1224–1229.

Ritsma, R. J. (**1963**). ''Existence region of the tonal residue. II.,'' J. Acoust. Soc. Am. **35**, 1241–1245.

Ruggero, M. A., Rich, N. C., Robles, L., and Recio, A. (**1996**). ''The effects of acoustic trauma, other cochlea injury and death on basilar membrane responses to sound,'' in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Axelsson, H. Borchgrevink, R. P. Hamernik, P. A. Hellstrom, D. Henderson, and R. J. Salvi (Thieme, Stockholm).

Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.

Shamma, S., and Klein, D. (**2000**). ''The case of the missing pitch templates: How harmonic templates emerge in the early auditory system,'' J. Acoust. Soc. Am. **107**, 2631–2644.

Siebert, W. M. (**1968**). ''Stimulus transformations in the peripheral auditory system,'' in *Recognizing Patterns*, edited by P. A. Kolers and M. Eden (MIT Press, Cambridge, MA).

Siebert, W. M. (**1970**). ''Frequency discrimination in the auditory system: Place or periodicity mechanisms,'' Proc. IEEE **58**, 723–730.

Srulovicz, P., and Goldstein, J. L. (**1983**). ''A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.

Turner, C. W., Burns, E. M., and Nelson, D. A. (**1983**). ''Pure tone pitch perception and low-frequency hearing loss,'' J. Acoust. Soc. Am. **73**, 966–975.

von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).

Ward, W. D. (**1954**). ''Subjective musical pitch,'' J. Acoust. Soc. Am. **26**, 369–380.

Woolf, N. K., Ryan, A. F., and Bone, R. C. (**1981**). ''Neural phase-locking properties in the absence of outer hair cells,'' Hear. Res. **4**, 335–346.

Zwislocki, J. (**1953**). ''Acoustic attenuation between the ears,'' J. Acoust. Soc. Am. **25**, 752–759.

Equal sensation curves for whole-body vibration expressed as a function of driving force

Neil J. Mansfield[a)]
Department of Human Sciences, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom

Setsuo Maeda[b)]
Department of Human Engineering, National Institute of Industrial Health, 21-1, Nagao 6-Chome, Tama-Ku, Kawasaki 214-8585, Japan

(Received 8 September 2004; revised 9 March 2005; accepted 10 March 2005)

Previous studies have shown that the seated human is most sensitive to whole-body vertical vibration at about 5 Hz. Similarly, the body shows an apparent mass resonance at about 5 Hz. Considering these similarities between the biomechanical and subjective responses, it was hypothesized that, at low frequencies, subjective ratings of whole-body vibration might be directly proportional to the driving force. Twelve male subjects participated in a laboratory experiment where subjects sat on a rigid seat mounted on a shaker. The magnitude of a test stimulus was adjusted such that the subjective intensity could be matched to a reference stimulus, using a modified Bruceton test protocol. The sinusoidal reference stimulus was 8-Hz vibration with a magnitude of 0.5 m/s^2 rms (or 0.25 m/s^2 rms for the 1-Hz test); the sinusoidal test stimuli had frequencies of 1, 2, 4, 16, and 32 Hz. Equal sensation contours in terms of seat acceleration showed data similar to those in the literature. Equal sensation contours in terms of force showed a nominally linear response at 1, 2, and 4 Hz, but an increasing sensitivity at higher frequencies. This is in agreement with a model derived from published subjective and objective fitted data. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1898703]

PACS numbers: 43.66.Wv, 43.40.Ng [RAL] Pages: 3853–3859

I. INTRODUCTION

Experimental work since the 1960s has indicated that perception of whole-body vibration is a function of vibration frequency (e.g., Miwa, 1969; Mansfield, 2004). Many studies have shown that the seated human is most sensitive to vertical vibration at about 4–5 Hz and most sensitive to horizontal vibration at lower frequencies (e.g., Parsons and Griffin, 1988). These data have been collated into a series of frequency weightings, such as those published in ISO2631-1 (1997) and BS6841 (1987) (e.g., Fig. 1). These weightings model the subjective response to vibration such that, for single-axis vibration

$$I(f) \propto a_{\text{weighted}}(f),$$

where $I(f)$ is the subjective rating of vibration intensity and $a_{\text{weighted}}(f)$ is the frequency-weighted acceleration at frequency f, although the relationship in terms of vibration magnitude is not necessarily linear (e.g., Morioka and Griffin, 2000).

The biomechanical response of the seated person has previously been assessed using kinetic (i.e., impedance and apparent mass) and kinematic (i.e., transmissibility) approaches. The apparent mass is calculated as the ratio of the force to the acceleration in the frequency domain

$$M(f) = \frac{F_{\text{seat}}(f)}{a_{\text{seat}}(f)},$$

where $M(f)$ is the apparent mass, $F_{\text{seat}}(f)$ is the force, and $a_{\text{seat}}(f)$ is the acceleration at frequency f at the seat surface. There is a peak in the vertical apparent mass (or driving point mechanical impedance) of the seated human at about 5 Hz that reduces in frequency as the magnitude of the vibration increases (e.g., Fairley and Griffin, 1989; Hinz and Seidel, 1987; Mansfield and Griffin, 2000; Smith, 1994). The general shape of the peak is not dependent on vibration waveform or subject age, if the effects of subject mass are taken into account (Giacomin, 2004; Mansfield and Maeda, 2004). In the horizontal direction, a similar nonlinearity has been observed, but with peaks in response at lower frequencies (Mansfield and Lundström, 1999a). The response of lumped parameter models (Mansfield and Lundström, 1999b) of the apparent mass of the seated human to whole-body vibration is shown in Fig. 2.

Considering the models of human responses to vibration depicted in Figs. 1 and 2, it appears that subjective ratings of vibration and biomechanical responses might be related. For vertical motion, both subjective and biomechanical responses peak at about 5 Hz. Similarly, for horizontal motion, both subjective and biomechanical responses peak at about 2 Hz. However, for vertical vibration, the roll-off at high frequencies for these models (and indeed for subject data) is not as great for the frequency weighting.

The similarity between the subjective and biomechanical models at low frequency is sufficient to question whether

[a)]Electronic mail: n.j.mansfield@lboro.ac.uk
[b)]Electronic mail: maeda@niih.go.jp

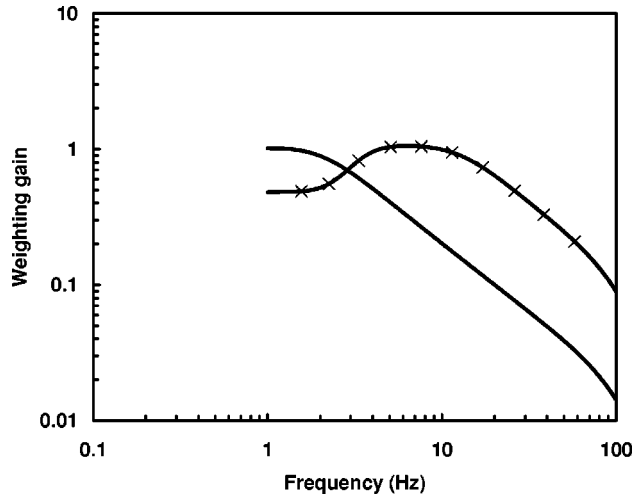


FIG. 1. Modulus of frequency weightings for vertical vibration (W_k, $-\times-$) and horizontal vibration (W_d, ———) as used in ISO2631-1 (1997).

there is an underlying link between the mechanical response (i.e., apparent mass) of the sitting person and their subjective response such that the apparent mass could effectively be used as a frequency weighting such that

$$I(f) \propto M(f) \times a_{\text{seat}}(f).$$

If this is the case, then considering the definition of apparent mass

$$I(f) \propto F_{\text{seat}}(f).$$

Therefore, subjective ratings of vibration intensity could be predicted solely from measurements of force at the seat. It should be noted that this technique would mean that no frequency weighting would be required for subjective assessments of vibration, and therefore it would be possible that interindividual differences could be accounted for as an inherent property of the method. Furthermore, it would mean that nonlinearities in response with respect to waveform and magnitude could be accounted for without the requirement for complex mathematical models. Although this approach is attractive, it is unclear whether it is viable either in terms of transducer technologies or whether the frequency range over which it might be effective is too restrictive. For example, the discrepancy between the apparent mass model and subjective response model could be less than or greater than that observed for individual subjects rendering the method more or less viable, respectively.

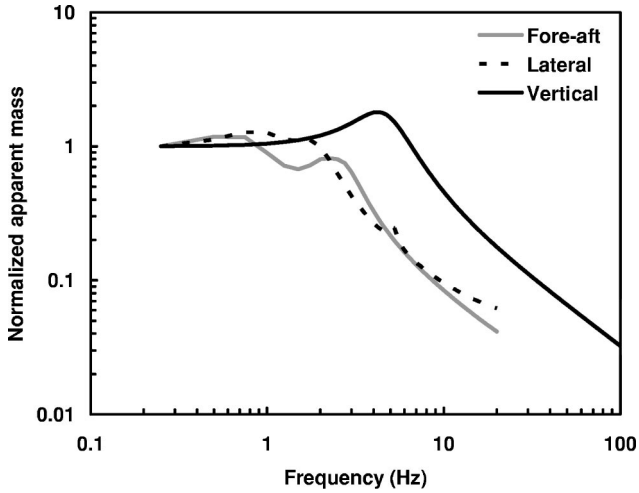


FIG. 2. Normalized responses of lumped parameter apparent mass models defined in Mansfield and Lundström, 1999a.

All known studies of human responses to vibration in the literature have scaled the dependent variable (biomechanical or subjective) in terms of the acceleration, velocity, or displacement. To investigate whether human response to whole-body vibration is proportional to unweighted force, experimental data are required that link subjective ratings directly to driving force.

This paper reports a study that had the following objectives:

(1) To define an equal comfort contour for whole-body vibration expressed in terms of force at the seat;
(2) To identify the frequency range (if any) over which subjective rating is not a function of frequency when expressed in terms of force; and
(3) To provide a preliminary evaluation of the feasibility of using measurements of force at the seat as a predictor for subjective ratings of whole-body vibration.

II. METHODS

A subjective intensity-matching protocol was used in the experiment. Participants were exposed to five groups of trials. Each trial used a pair of sinusoidal vertical vibration stimuli: a reference stimulus at 8 Hz and a test stimulus with a frequency of either 1, 2, 4, 16, or 32 Hz. Eight Hz was used for the reference to ensure that the reference and test stimuli frequencies were as close as possible when considered on a logarithmic scale (i.e., a maximum of 3 octaves), while avoiding the peak in apparent mass that could occur at 4 Hz for some subjects. Each stimulus had a duration of 4 s consisting of a 0.5-s taper, 3 s of steady sinusoidal vibration, and finishing with a 0.5-s taper. There was a pause of 1 s between the reference and the test stimuli. For the 1-Hz trials, the reference stimulus (i.e., the 8-Hz stimulus) had a vibration magnitude of 0.25 ms^{-2}; for all other trials, the reference stimulus had a vibration magnitude of 0.5 ms^{-2}. The lower magnitude at 1 Hz was required to ensure that the experimental conditions remained within the specification of the vibrator.

Each trial used a modified Bruceton test protocol. After the pair of stimuli, subjects judged whether the first (reference) or second (test) stimulus had the greater intensity, or whether they felt the same. If the reference was perceived as being more intense, the magnitude of the test vibration was increased by 20%. If the reference was perceived as being less intense, the magnitude of the test vibration was reduced by 20%. A step size of 20% was used, as this is approximately two times the difference threshold for whole-body vibration perception (e.g., Mansfield and Griffin, 2000; Morioka and Griffin, 2000). If the two stimuli were rated as being equal, then the magnitude of the test vibration was scaled by 20% in the same direction as the previous change. Therefore, the vibration magnitude would continually step up or down between a series of reversal points. The proce-

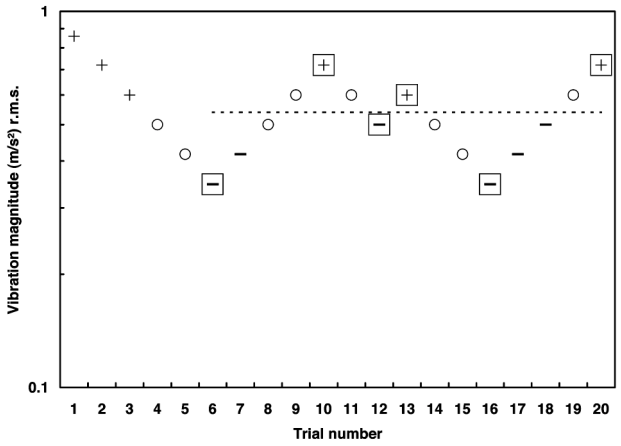


FIG. 3. Example of experimental protocol for one group of trials. After each rating of the test stimulus being more intense than the reference (+), the test vibration magnitude decreased by 20%. After each rating of the test stimulus being less intense than the reference (−), the test vibration magnitude increased by 20%. If the stimuli were rated as equal (○), the test magnitude was stepped up or down, repeating the previous change. The arithmetic mean of the six reversals (□) was used to determine the magnitude of equal sensation for the reference and test stimuli for that group of trials (---).

dure was continued until six reversals had been completed (Fig. 3). The value used for the matched intensity in the data analysis was the arithmetic mean of the six reversals for the acceleration and for the force. The starting magnitude for the test stimuli was randomly set at either 0.29 or 0.86 ms^{-2} rms, W_k weighted, apart from for the 1-Hz stimulus where these were set at 0.14 or 0.43 ms^{-2} rms Therefore, the first test stimulus should have been clearly perceived as either lower or higher than the reference stimulus, respectively, assuming that the W_k weighting is a reasonable model of whole-body vibration perception. The frequency weighting was only used to calculate the starting magnitude for each group of trials: all other reports of vibration magnitude in this paper are unweighted. The trial groups were presented in a balanced random order.

Twelve male subjects participated in the experiment and were instructed to sit in a comfortable, relaxed upright posture. The subjects had a mean age of 27.9 years (s.d. 8.6), a mean weight of 65.8 kg (s.d. 11.6), and a mean stature of 171 cm (s.d. 11.6).

An electro-hydraulic vibrator with low distortion was used to generate the vibration stimuli. This system had a peak-to-peak stroke of 100 mm and was situated in the laboratories of the National Institute of Industrial Health, Kawasaki, Japan. Acceleration at the seat was measured using a Brüel & Kjær 4370 accelerometer. The force at the seat was measured using a Kistler 9286A force plate, which acted as the seat surface. The influence of the mass of the plate (approximately 15 kg) was removed using a mass cancellation technique in the frequency domain. No backrest was used in the experiment. The vibrator was controlled by, and force and acceleration signals were acquired by, an HVLab data acquisition system at 512 samples per second via antialiasing filters set at 170 Hz. Calculations were based on acquired acceleration magnitudes for the trial stimuli, rather than ''desired'' vibration magnitudes. The control software included an algorithm to automatically equalize the next vibration stimulus based on the transfer function measured for the previous vibration stimulus. Measured vibration magnitudes remained within 5% of the desired magnitudes.

All subjects were explained the purpose of the study and gave their consent to participate. They were free to withdraw at any time, although all subjects completed the trials. The experiment was approved by the Research Ethic Committee of the National Institute of Industrial Health.

III. RESULTS

A full set of data was obtained for 11 of the 12 subjects. An equipment difficulty meant that data at 4 Hz were not obtained for one subject, although all other conditions were successfully measured. As ratings at 1 Hz were obtained relative to a reference with a magnitude of 50% of that used for all other conditions, measured values for acceleration and force were scaled by a factor of 2 to enable direct comparisons to be made.

For the acceleration data, most subjects showed that they were relatively less sensitive to the vibration at frequencies greater than the reference (i.e., test frequencies of 16 and 32 Hz). At frequencies lower than the reference, most subjects also showed a slight reduction in sensitivity, but this was not observed for all individuals (Fig. 4, Table I). Median data for the subjects expressed in terms of acceleration show that there was not a clear difference between the sensitivities at 1 and 2 Hz, sensitivities were slightly greater at 4 and 8 Hz (the reference frequency), and that there was a steady decrease in sensitivity from 8 Hz through to 32 Hz. The differences between the data obtained at 32 Hz and all other frequencies were significant ($p<0.005$, Wilcoxon). The only other significant difference occurred between the 2- and 8-Hz data ($p<0.05$), although differences approached significance between 1 and 8 Hz ($p=0.07$) and between 8 and 16 Hz ($p=0.08$).

Due to the differences in the subject masses, the forces for the reference stimuli varied, despite the acceleration magnitude being controlled. Data were therefore normalized by dividing by the force at the reference magnitude to enable comparison between subjects (Table II). For the force data, all subjects showed that they were more sensitive to higher frequency forces than lower frequency forces. At frequencies above the reference (8 Hz), all subjects showed an increase in sensitivity (Fig. 5). At frequencies below the reference, subjects tended to be slightly less sensitive in terms of force. All subjects were less sensitive to the vibration at 2 and 4 Hz when compared to 8 Hz, and ten of the subjects were less sensitive to vibration at 1 Hz when compared to 8 Hz (i.e., a greater force was required for parity). Most subjects showed a relatively equal response at 1, 2, and 4 Hz. The forces measured at 16 and 32 Hz were all significantly lower than data obtained at lower frequencies ($p<0.005$, Wilcoxon), although there was no significant difference between the ratings at 16 and 32 Hz. Similarly, significant differences were observed between the data at 8 and 1 Hz ($p<0.05$) and between 8 and 2 or 4 Hz ($p<0.005$). There were no significant differences between the results obtained at 1, 2, and 4 Hz.

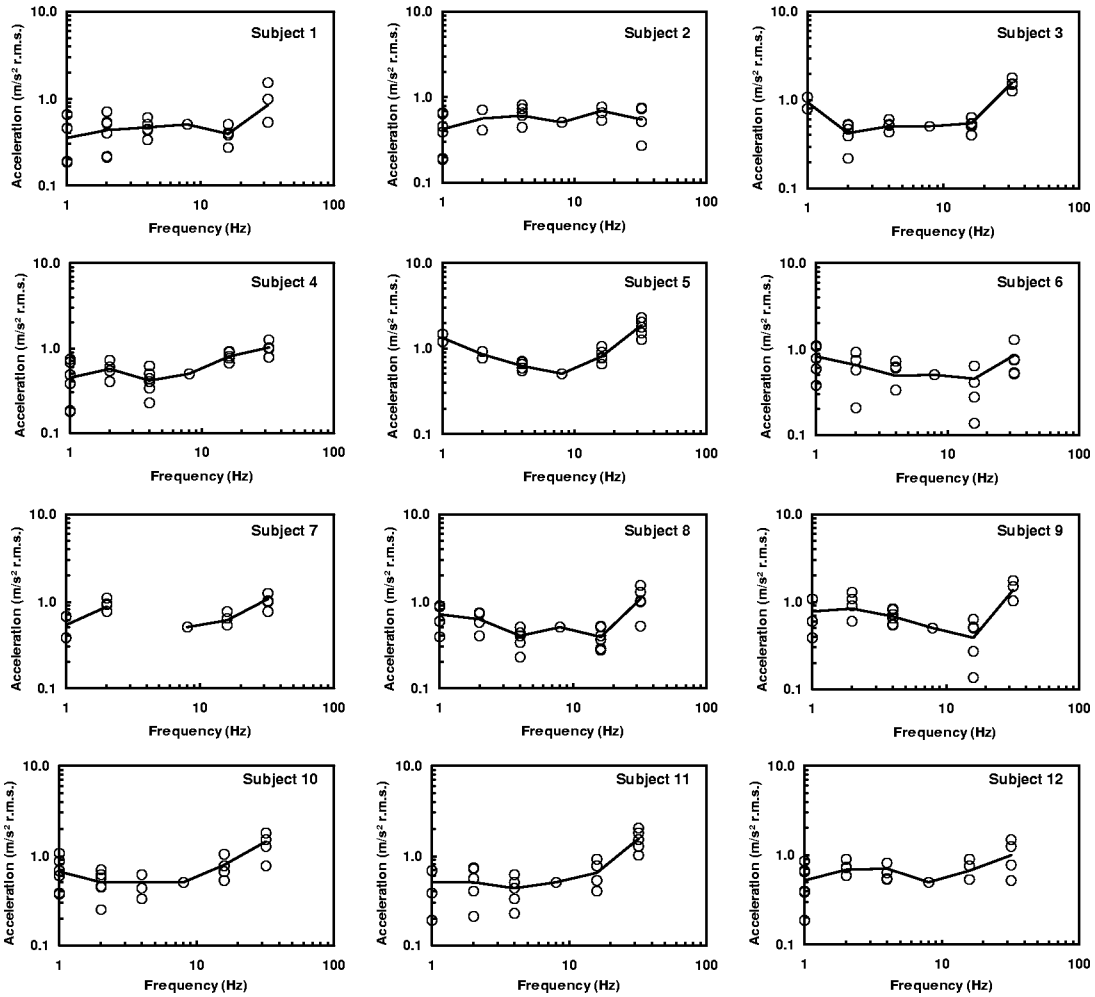


FIG. 4. Equal sensation curves for 12 subjects exposed to whole-body vertical vibration. Data represent accelerations corresponding to three positive and three negative reversals at each frequency (○) and mean of all reversals at each frequency (———).

IV. DISCUSSION

This study used a single reference frequency of 8 Hz. It is possible that the proximity of the reference and test frequencies could affect subjective responses; this could be investigated in future research. Unlike some previously published studies, the experiment did not use a binary forced-choice protocol but a modified Bruceton protocol as described in Sec. II. The reason for this was to maintain the

TABLE I. Individual data corresponding to equal sensation contours for rating of whole-body vibration in terms of acceleration.

	Acceleration (m/s^2 rms)					
Frequency	1 Hz	2 Hz	4 Hz	8 Hz	16 Hz	32 Hz
Subject 1	0.36	0.43	0.47	0.50	0.39	0.85
Subject 2	0.42	0.56	0.61	0.50	0.69	0.55
Subject 3	0.94	0.42	0.51	0.50	0.54	1.56
Subject 4	0.44	0.56	0.42	0.50	0.81	1.01
Subject 5	1.35	0.84	0.63	0.50	0.81	1.86
Subject 6	0.83	0.66	0.49	0.50	0.46	0.85
Subject 7	0.53	0.88	⋯	0.50	0.61	1.05
Subject 8	0.70	0.63	0.40	0.50	0.39	1.06
Subject 9	0.77	0.84	0.68	0.50	0.39	1.35
Subject 10	0.66	0.51	0.51	0.50	0.78	1.44
Subject 11	0.50	0.50	0.44	0.50	0.65	1.56
Subject 12	0.52	0.69	0.70	0.50	0.67	1.01
25th percentile	0.49	0.51	0.45	0.50	0.44	0.97
Median	0.60	0.60	0.51	0.50	0.63	1.05
75th percentile	0.78	0.73	0.62	0.50	0.72	1.47

TABLE II. Individual data corresponding to equal sensation contours for rating of whole-body vibration in terms of rms normalized force.

	Normalized force (rms)					
Frequency	1 Hz	2 Hz	4 Hz	8 Hz	16 Hz	32 Hz
Subject 1	0.855	1.102	1.387	1.000	0.338	0.168
Subject 2	0.717	1.014	1.267	1.000	0.449	0.094
Subject 3	2.856	1.382	2.133	1.000	0.437	0.632
Subject 4	1.351	1.829	1.788	1.000	0.442	0.410
Subject 5	5.118	3.606	4.551	1.000	0.509	0.658
Subject 6	3.162	2.689	2.755	1.000	0.361	0.467
Subject 7	2.119	3.879	⋯	1.000	0.676	0.379
Subject 8	1.772	1.702	1.196	1.000	0.436	0.150
Subject 9	2.036	2.361	2.293	1.000	0.372	0.479
Subject 10	1.568	1.206	1.593	1.000	0.330	0.358
Subject 11	1.005	1.085	1.176	1.000	0.401	0.331
Subject 12	1.099	1.493	1.736	1.000	0.479	0.393
25th percentile	1.076	1.180	1.327	1.000	0.369	0.290
Median	1.670	1.598	1.736	1.000	0.436	0.386
75th percentile	2.303	2.443	2.213	1.000	0.456	0.470

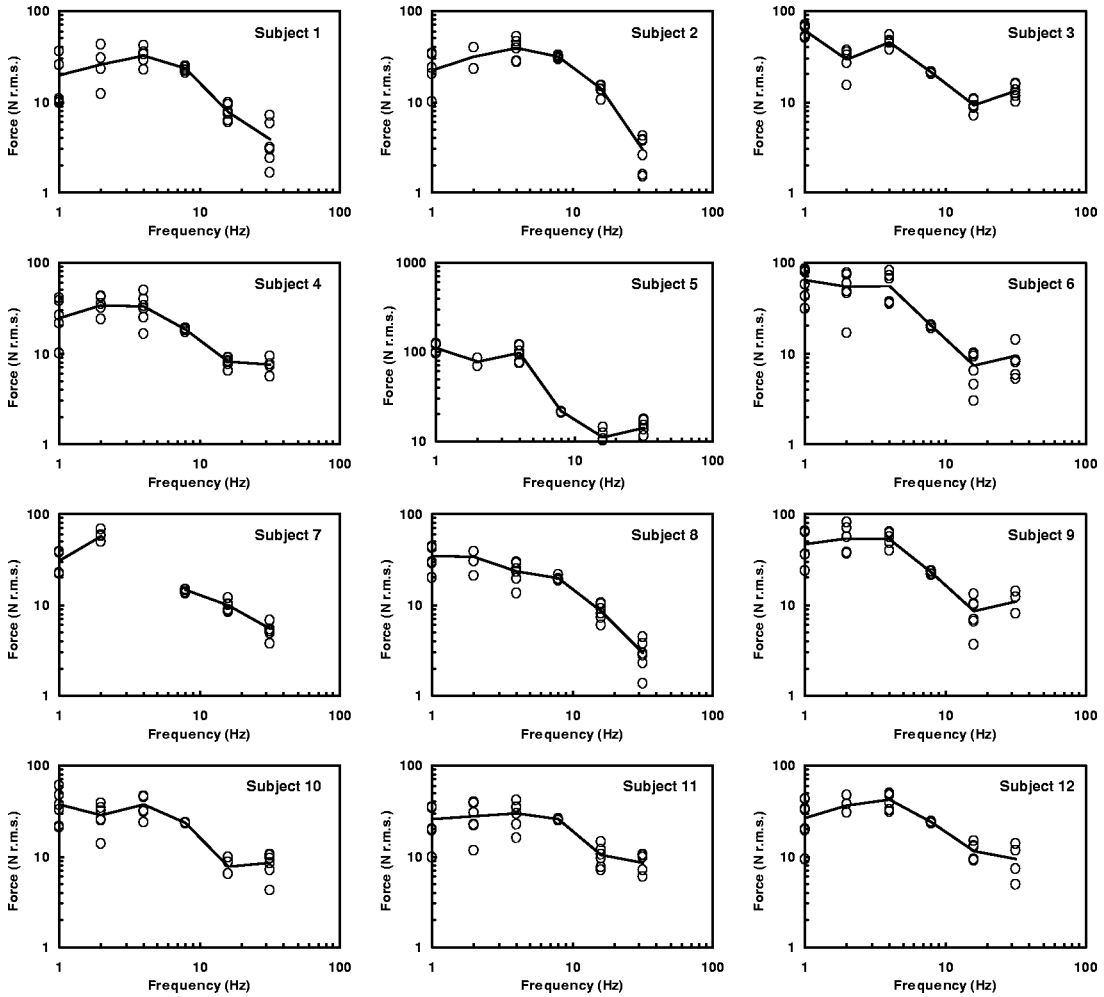


FIG. 5. Equal sensation curves for 12 subjects exposed to whole-body vertical vibration. Data represent the rms force corresponding to three positive and three negative reversals at each frequency (○) and mean of all reversals at each frequency (———).

confidence of the subjects while making their judgments. Anecdotes from subjects used by the authors in previous experiments have suggested that participants feel uncomfortable being forced to ''guess'' when performing forced-choice protocols, and it was suspected that this could have contributed to occasional loss of concentration of some subjects during previous studies. A further advantage of the protocol used here is that it does not rely on interpretation of probabilities of chance ''correct'' answers as occurs for forced-choice up-and-down method of limits techniques (Maeda and Griffin, 1995). The choice of experimental protocol could have affected the results, although this is likely to have been a systematic difference and therefore would not alter the relative ratings, as were sought in this experiment. Nevertheless, as results from this study are within the range of those in the literature, all indications are the methodology can be considered valid.

The equal sensation data in terms of acceleration produced in this experiment are comparable to other studies in the literature which have previously been identified as having minimal bias (Griffin *et al.*, 1982). If published data sets are scaled to give a value of 0.5 at 8 Hz, then the trends are similar between studies (Fig. 6). All show frequencies of greatest sensation to occur between 4 and 8 Hz, and a reduced sensation at higher and lower frequencies. The gradient of the reduction in sensation with increasing proximity from the 4–8-Hz range shows some variation between studies. Gradients obtained in this study were not as steep as some of those observed previously.

The most commonly used frequency weighting for

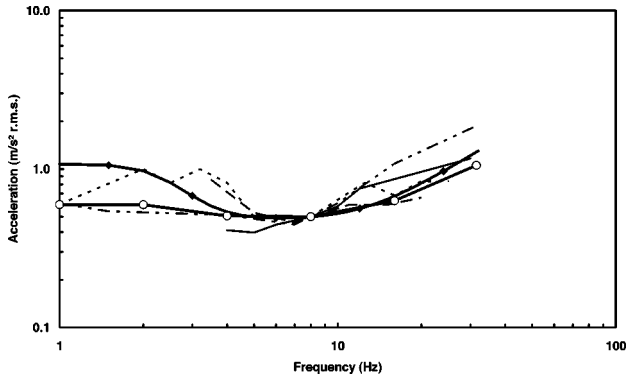


FIG. 6. Comparison of median equal sensation acceleration data obtained in this study (–○–) with data from Miwa (1967; – – – –), Shoenberger and Harris (1971; – – –), Jones and Saunders (1972; ———), Griffin *et al.* (1982; ---), and an inverted W_k frequency weighting (–♦–), all scaled to give a value of 0.5 at 8 Hz (the reference frequency in this study).

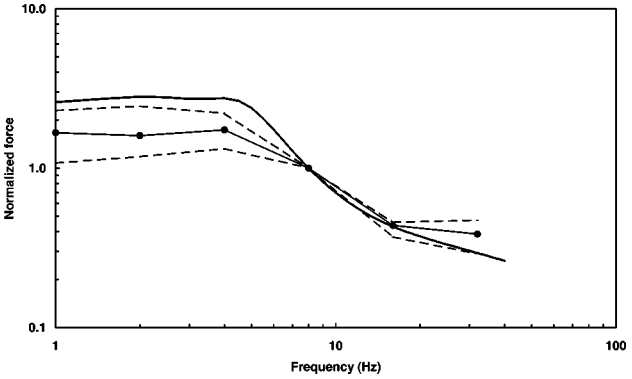


FIG. 7. Comparison of median equal sensation normalized force data obtained in this study (-●-) with interquartile ranges (---) and results from the model based on frequency weighting and biomechanical models in the literature (———), normalized at 8 Hz.

whole-body vertical vibration is the W_k weighting (ISO2631-1, 1997). At frequencies of 4 Hz and above, the results for acceleration in this study correspond closely to the W_k curve (Fig. 6). However, below 4 Hz the W_k curve shows a more rapid reduction in sensitivity than observed here. In agreement with the literature, the ratings of vibration intensity in this study had a discrepancy of approximately 2:1 at 1 Hz when compared to W_k. This indicates that W_k might underestimate subjective ratings of vibration magnitude at low frequency. However, W_k is also designed to be applicable for health risk assessments. Biomechanical data (e.g., Smith, 2000; Boileau *et al.*, 2002; Nawayseh and Griffin, 2003) tend to show a resonance in the response of the seated body to vibration at about 5 Hz. Therefore, data from this study alone do not necessarily imply that the frequency-weighting curve should be changed (at least for health risk assessments).

Previous studies have not generally reported subjective and biomechanical responses to whole-body vibration within the same experiment. Therefore, it is not possible to derive the relationship between any individual's subjective response to force and the frequency of the force. However, it is possible to use a model of subjective response (i.e., a frequency weighting) in combination with a model of biomechanical response (i.e., a lumped parameter apparent mass model) to derive a generalized expected profile for subjective responses with respect to force at any frequency. As apparent mass is defined as the ratio of the force to the acceleration at any frequency, and the frequency weighting is modeled to the subjective rating of acceleration at any frequency. Division of the apparent mass by the frequency weighting will result in a model of an equal sensation response in terms of force. A model derived from the W_k weighting and the lumped parameter model presented in Mansfield and Lundström (1999b), where the data sets are scaled to give a value of unity at 8 Hz, is shown in Fig. 7 and compared to the data obtained in this study. This shows that at frequencies below 5 Hz the effect of frequency observed in the weighting and in the apparent mass cancel one another out such that the subjective response is linear with respect to frequency. At frequencies above 5 Hz the equal sensation curve decays, indicating that a lower force is required for an equivalent sensation. Therefore, a greater sensation occurs for the same rms force.

The similarity between the shape of the response of the model of an equal sensation curve for force based on fitted curves and the results obtained in this study is clear. Although in the representation illustrated in Fig. 7 the differences between the curves occur at low frequency, this is due to the normalization occurring at 8 Hz; if the normalization occurred at, for example, 1 Hz, the difference would be observed at the higher frequencies. These data therefore indicate that the gradient of increasing sensation at high frequencies is steeper in the model than that observed in this study. The frequency at which the response starts to be affected by vibration frequency was similar for the model and the experimental data.

Although subjects were not interviewed in a structured manner following the trials, it was remarked that the low-frequency stimuli were perceived by the whole body, whereas the sensations of high-frequency vibration were localized to the skin and muscle of the thigh and buttocks. This corresponds to published data concerning transmission of vibration through the seated person which show that sites remote from the driving point are effectively isolated from the vibration at higher frequencies (e.g., Paddan and Griffin, 1998). Therefore, the mechanism for the increased sensitivity at high frequencies must either be due to the nature of the underlying physiology (e.g., fast-adapting nerve endings in the skin sensitive to vibration above 10 Hz) or an indirect result of the biomechanical response. The increased sensitivity to force at high frequencies being due to the biomechanical response is the most plausible, as less force is required to generate the same acceleration at high frequencies when compared to low frequencies, as can be seen in the apparent mass of the seated body (Fig. 2).

This experiment has indicated that, for low-frequency vibration, an assessment methodology based on driving force could be envisaged. However, the frequency range over which this would be of value would be constrained to frequencies below about 5 Hz. Furthermore, there are practical difficulties in measuring the driving force in vehicles, although some studies have reported such data (e.g., Holmlund, 1999). Considering that most environments where individuals are exposed to vibration have a significant proportion of vibration energy at frequencies above this 5-Hz threshold, the use of force cannot be considered a pragmatic method of vibration assessment.

V. CONCLUSIONS

This study has shown that the sensitivity of seated humans to whole-body vibration is greatest for acceleration at frequencies between 4 and 8 Hz. This is in agreement with the literature, although a slightly different psychophysical technique was used. Subjective ratings of vibration in terms of force showed that the greatest sensitivity occurred at high frequencies. A nominally linear response occurred between 1 and 4 Hz. This is in agreement with a model of response based on a frequency weighting and apparent mass model in the literature.

ACKNOWLEDGMENTS

The authors acknowledge the assistance of staff and students at Tokyo Metropolitan University.

Boileau, P. E., Rakheja, S., and Wu, X. (**2002**). ''A body mass-dependent mechanical impedance model for applications in vibration seat testing,'' J. Sound Vib. **253**(1), 243–264.

British Standards Institution (**1987**). ''Measurement and evaluation of human exposure to whole-body mechanical vibration and repeated shock,'' BS6841, BSI, London.

Fairley, T. E., and Griffin, M. J. (**1989**). ''The apparent mass of the seated human body: Vertical vibration,'' J. Biomech. **22**(2), 81–94.

Giacomin, J. (**2004**). ''Apparent mass of small children: Experimental measurements,'' Ergonomics **47**(13), 1454–1474.

Griffin, M. J., Parsons, K. C., and Whitham, E. M. (**1982**). ''Vibration and comfort. IV. Application of experimental results,'' Ergonomics **25**(8), 721–739.

Hinz, B., and Seidel, H. (**1987**). ''The nonlinearity of the human body's dynamic response during sinusoidal whole body vibration,'' Ind. Health **25**, 169–181.

Holmlund, P. (**1999**). ''Absorbed power and mechanical impedance of the seated human measured within a real vehicle environment compared with single axis laboratory data,'' J. Low Freq. Noise Vib. Active Control **18**(3), 97–110.

International Organization for Standardization (**1997**). ''Mechanical vibration and shock—evaluation of human exposure to whole-body vibration: ISO2631-1,'' ISO, Geneva.

Jones, A. J., and Saunders, D. J. (**1972**). ''Equal comfort contours for whole-body vertical pulsed sinusoidal vibration,'' J. Sound Vib. **23**, 1–14.

Maeda, S., and Griffin, M. J. (**1995**). ''A comparison of vibrotactile thresholds on the finger obtained with different measuring algorithms,'' in Proceedings of Stockholm Workshop 1994: Hand–Arm Vibration: Diagnostics and Quantitative Relationships to Exposure, Arbete och Halsa, 1995.

Mansfield, N. J. (**2004**). *Human Response to Vibration* (CRC Press, Boca Raton, FL).

Mansfield, N. J., and Griffin, M. J. (**2000**). ''Difference thresholds for automobile seat vibration,'' Appl. Ergon **31**(3), 255–261.

Mansfield, N. J., and Griffin, M. J. (**2000**). ''Nonlinearities in apparent mass and transmissibility during exposure to whole-body vibration,'' J. Biomech. **33**(8), 933–941.

Mansfield, N. J., and Lundström, R. (**1999a**). ''The apparent mass of the human body exposed to nonorthogonal horizontal vibration,'' J. Biomech. **32**(12), 1269–1278.

Mansfield, N. J., and Lundström, R. (**1999b**). ''Models of the apparent mass of the seated human body exposed to horizontal whole-body vibration,'' Aviat., Space Environ. Med. **70**(12), 1166–1172.

Mansfield, N. J., and Maeda, S. (**2005**). ''Comparison of the apparent mass of the seated human measured using random and sinusoidal vibration,'' Ind. Health **43**(1), 233–240.

Miwa, T. (**1967**). ''Evaluation methods for vibration effect. 1. Measurements of threshold and equal sensation contours of whole-body for vertical and horizontal vibrations,'' Ind. Health **5**, 183–205.

Miwa, T. (**1969**). ''Evaluation methods for vibration effect. Part 8: The vibration greatness of random waves,'' Ind. Health **7**, 89–115.

Morioka, M., and Griffin, M. J. (**2000**). ''Difference thresholds for intensity perception of whole-body vertical vibration: Effect of frequency and magnitude,'' J. Acoust. Soc. Am. **107**(1), 620–624.

Nawayseh, N., and Griffin, M. J. (**2003**). ''Nonlinear dual-axis biodynamic response to vertical whole-body vibration,'' J. Sound Vib. **268**(3), 503–523.

Paddan, G. S., and Griffin, M. J. (**1998**). ''A review of the transmission of translational seat vibration to the head,'' J. Sound Vib. **215**(4), 863–882.

Parsons, K. C., and Griffin, M. J. (**1988**). ''Whole-body vibration perception thresholds,'' J. Sound Vib. **121**(2), 237–258.

Shoenberger, R. W., and Harris, C. S. (**1971**). ''Psychophysical assessment of whole-body vibration,'' Hum. Factors **13**, 41–50.

Smith, S. D. (**1994**). ''Nonlinear resonance behavior in the human exposed to whole-body vibration,'' Shock Vib. **1**(5), 439–450.

Smith, S. D. (**2000**). ''Modeling differences in the vibration response characteristics of the human body,'' J. Biomech. **33**(11), 1513–1516.

Pitch discrimination interference: The role of pitch pulse asynchrony

Hedwig Gockel[a] and Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 27 January 2005; revised 4 March 2005; accepted 4 March 2005)

Gockel, Carlyon, and Plack [J. Acoust. Soc. Am. **116**, 1092–1104 (2004)] showed that discrimination of the fundamental frequency ($F0$) of a target tone containing only unresolved harmonics was impaired when an interfering complex tone with fixed $F0$ was added to the target, but filtered into a lower frequency region. This pitch discrimination interference (PDI) was greater when the interferer contained resolved harmonics than when it contained only unresolved harmonics. Here, it is examined whether this occurred because, when the interferer contained unresolved harmonics, ''pitch pulse asynchrony (PPA)'' between the target and interferer provided a cue that enhanced performance; this was possible in the earlier experiment because both target and interferer had components added in sine phase. In experiment 1, it was shown that subjects were moderately sensitive to the direction of PPA across frequency regions. In experiments 2 and 3, PPA cues were eliminated by adding the components of the target only, or of both target and interferer, in random phase. For both experiments, an interferer containing resolved harmonics produced more PDI than an interferer containing unresolved harmonics. These results show that PDI is smaller for an interferer with unresolved harmonics even when cues related to PPA are eliminated. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1898084]

PACS numbers: 43.66.−x, 43.66.Fe, 43.66.Hg, 43.66.Ba [JHG] Pages: 3860–3866

I. INTRODUCTION

Most models for the pitch perception of complex tones are based on the assumption that information can be combined across different frequency regions in order to estimate the fundamental frequency ($F0$) of a complex tone (Goldstein, 1973; Terhardt, 1974; Moore, 1982; Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). This is consistent with empirical evidence showing that information can be combined across frequency regions (Moore *et al.*, 1984). However, it has also commonly been assumed that it is possible to listen selectively to specific frequency regions in order to extract the $F0$s of multiple complex tones presented simultaneously. For example, it has been shown that the ability to identify two vowels presented simultaneously is better when the two vowels have different $F0$s than when they have the same $F0$ (Scheffers, 1983). Models to explain this effect (Summerfield and Assmann, 1987; Assmann and Summerfield, 1990; 1994; Meddis and Hewitt, 1992; de Cheveigné, 1993) are based on the fact that the two vowels have formants at different frequencies, so that the spectrum is dominated at some frequencies by one vowel and at other frequencies by the other vowel. The models all assume that information can be extracted independently from different regions of the basilar membrane.

Recently, Gockel *et al.* (2004) showed that listeners are not always able to extract the $F0$ of harmonics in one frequency region independently of the $F0$ of harmonics in another frequency region. In other words, they demonstrated an across-frequency interference effect in pitch perception, which they called ''pitch discrimination interference'' (PDI). They measured sensitivity (d') for $F0$ discrimination between two sequentially presented complex (target) tones containing only high, unresolved harmonics. In their first experiment, the target was filtered between 1375 and 15 000 Hz and had a nominal $F0$ of 88 Hz. The target was presented either alone or with an additional complex tone, referred to as the ''interferer.'' The interferer was filtered between 125 and 625 Hz, and its $F0$ varied between 88 and 114.4 Hz across blocks. Sensitivity was significantly reduced in the presence of the interferer, and this PDI effect decreased as the interferer's $F0$ was moved further above that of the target, i.e., the PDI effect was tuned for $F0$.

In a second experiment, Gockel *et al.* (2004) showed that increasing the level of a synchronously gated low-pass noise that spectrally overlapped with the interferer reduced the PDI effect, suggesting that the tonality of the interferer has an important influence on PDI.

In a third experiment, Gockel *et al.* (2004) compared the magnitude of PDI for interferers with resolved and with unresolved harmonics. The target was filtered between 3900 and 5400 Hz and had a nominal $F0$ of either 88 or 250 Hz. It was presented either alone or with an interferer filtered between 1375 and 1875 Hz and with an $F0$ corresponding to the nominal $F0$ of the target. PDI was larger in the presence of the interferer with resolved harmonics (250 Hz $F0$) than in the presence of the interferer with unresolved harmonics (88-Hz $F0$).

[a]Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

 0001-4966/2005/117(6)/3860/7/$22.50

In a final experiment, Gockel *et al.* (2004) investigated the role of perceptual grouping processes in PDI. They showed that PDI was reduced but not eliminated when the interferer was gated on 200 ms before and off 200 ms after the target, and that some PDI was observed with a continuous interferer. These findings suggest that the magnitude of PDI is influenced by perceptual grouping processes, but it occurs even under conditions of onset and offset asynchrony of the target and interferer, which would usually be assumed to lead to clear perceptual segregation (Bregman, 1990; Darwin and Carlyon, 1995).

The present paper is especially concerned with Gockel *et al.*'s finding that PDI was greater when the interferer contained resolved harmonics than when it contained unresolved harmonics. This finding has implications for the question of whether there are two mechanisms for determining the pitch of complex tones, one for resolved harmonics and the other for unresolved harmonics (Carlyon and Shackleton, 1994; Meddis and O'Mard, 1997; Carlyon, 1998; Grimault *et al.*, 2002; Micheyl and Oxenham, 2004). If there are two pitch mechanisms whose outputs can be assessed independently, one might expect greater PDI when the interferer and target have pitches determined by the same mechanism than when their pitches are determined by different mechanisms. Since the target contained only unresolved harmonics, this leads to the prediction that the interferer containing only unresolved harmonics would produce more PDI than the target containing mainly resolved harmonics. In fact, as already noted, the opposite was found.

Gockel *et al.* (2004) suggested that the magnitude of PDI is partly determined by the salience of the pitch of the interferer. Complex tones containing resolved harmonics have a clearer and more salient pitch than complex tones containing unresolved harmonics (Plomp, 1967; Ritsma, 1967; Moore and Rosen, 1979; Moore and Glasberg, 1988; Moore and Peters, 1992; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994), and this might be the cause of the greater PDI produced by the former. We consider here an alternative explanation, that, when the interferer contained only unresolved harmonics, there was a cue available that enhanced the discrimination of the $F0$ of the target, which partially offset the PDI effect; this cue was not available when the interferer with resolved harmonics was used.

The cue in question arises from the fact that complex tones containing unresolved harmonics evoke waveforms on the basilar membrane with a periodicity equal to the $F0$ of the complex. The peak factor of these waveforms depends on the phases with which the components are added. In experiment 3A of Gockel *et al.* (2004), both the target and the interferer had components added in sine phase, which leads to waveforms with high peak factors on the basilar membrane, when the harmonics are unresolved (Rhode and Recio, 2001; Moore, 2002). The waveforms have a single major peak, sometimes called a ''pitch pulse,'' per period. Human listeners have been shown to be sensitive to asynchronies of pitch pulses in different frequency regions (Patterson, 1987; Summerfield and Assmann, 1991; Carlyon, 1994; Carlyon and Shackleton, 1994). When two complex tones are presented simultaneously in different frequency regions, and when they have different $F0$s, the pitch pulse asynchrony (PPA) continuously changes, and listeners can use this as a cue for detecting the difference in $F0$ (Carlyon, 1994). In experiment 3A of Gockel *et al.* (2004), subjects may have been able to use PPA as a cue for discriminating the $F0$ of the target when the interferer contained unresolved harmonics; this would depend on subjects being able to judge the direction or pattern of change of the PPA (a more detailed explanation is provided later in the paper). Alternatively, or in addition, PPA might have provided a cue allowing perceptual segregation of the target and interferer (Summerfield and Assmann, 1991), thus reducing the PDI. However, this cue would not have been available when the interferer contained resolved harmonics, as complex tones with resolved harmonics do not lead to waveforms on the basilar membrane with high peak factors.

In the experiments reported here, we assessed whether the use of PPA cues could account for the finding of Gockel *et al.* (2004) that PDI was greater when the interferer contained resolved harmonics than when it contained unresolved harmonics. Experiment 1 showed that subjects were moderately sensitive to the direction of PPA across frequency regions, making it plausible that PPA cues might have been used in experiment 3A of Gockel *et al.* (2004) to improve performance and thereby to offset the PDI produced by an interferer with unresolved harmonics. Experiments 2 and 3 replicated experiment 3A of Gockel *et al.* (2004), except for the choice of starting phases of the components in the complex tones. In experiment 2, the components of the target were added in random phase, while those of the interferer were added in sine phase. In experiment 3, the components of both the target and the interferer were added in random phase. The use of random phase eliminates or substantially reduces any PPA cues, since distinct waveform peaks on the basilar membrane do not occur for random-phase harmonic complexes (Alcántara *et al.*, 2003).

II. EXPERIMENT 1: SENSITIVITY TO THE DIRECTION OF PPA

A. Rationale

In experiment 3A of Gockel *et al.* (2004), the components of both the target and interferer were added in sine phase (starting phase of 0°), which meant that, at the start of each stimulus, the pitch pulses of the target and interferer were synchronous. When the interferer contained only unresolved harmonics, the $F0$ of the interferer was fixed at 88 Hz. The $F0$ of the target differed across the two intervals of a forced-choice trial by 3.5% (3.08 Hz), so the $F0$ was 86.46 Hz in one interval and 89.54 Hz in the other interval. If the target and interferer were played continuously, the PPA between them would go through one complete cycle in 1/1.54 s, i.e., 649 ms. In other words, the pitch pulses would be synchronous again after 649 ms. In fact, the stimulus duration was 400 ms, so the PPA went through less than one cycle. When the $F0$ of the target was higher than that of the interferer, the pitch pulses of the target were initially advanced in time relative to those of the interferer, until, after 325 ms, the pitch pulses of the target were midway between

those of the interferer, leading to the maximum possible PPA of about 5.7 ms. When the $F0$ of the target was lower than that of the interferer, the pitch pulses of the target were initially delayed in time relative to those of the interferer by corresponding amounts. Thus, if subjects could discriminate whether the pitch pulses of the target were leading or lagging those of the interferer during the first 325 ms of the stimulus, they could use this as a cue to determine the direction of the $F0$ change of the target.

It has previously been shown that listeners can discriminate a stimulus with a PPA from a stimulus without a PPA (Carlyon, 1994). Also, Summerfield and Assmann (1991) showed that, for a very low $F0$, listeners could use PPA as a cue to enhance the detection of one vowel in the presence of another (three out of three listeners) and to enhance the identification of two simultaneously presented vowels (one out of three listeners). This benefit of PPA was observed when the common $F0$ of the two vowels was 50 Hz, but not when it was 100 Hz. However, in their experiment the PPA was associated with an onset and offset asynchrony, so it is not clear whether the observed effects were due to PPA alone. To our knowledge it has not been determined whether subjects are sensitive to the *direction* of PPA, i.e., whether they can distinguish whether the pitch pulses in one frequency region are leading or lagging those in another frequency region. Such an ability would be necessary for the subjects of Gockel *et al.* (2004) to have been able to use the PPA as a cue for discriminating the $F0$ of the target. In the present experiment, the ability to discriminate the direction of PPA was assessed using stimuli very similar to those of Gockel *et al.* (2004).

B. Stimuli

Subjects had to discriminate differences in the direction of PPA between two sequentially presented complex tones with an $F0$ of 88 Hz. Each tone was composed of two parts, one corresponding to the target tone and one to the interferer, as used by Gockel *et al.* (2004). We use the same names here, even though the task was to discriminate the direction of PPA between the target and interferer. The target was bandpass filtered between 3900 and 5400 Hz (3-dB down points, slopes of 48 dB/oct). The interferer was bandpass filtered between 1375 and 1875 Hz (3-dB down points, slopes of 48 dB/oct). Both the target and the interferer contained only unresolved harmonics (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). The level per component was always 45 dB SPL. To mask possible distortion products, a continuous pink background noise was presented with a spectrum level of 15 dB (re: 20 μPa) at 1 kHz. The nominal stimulus duration was 400 ms, including 5-ms raised-cosine onset and offset ramps. The interval between the two stimuli within a trial was 500 ms.

In a given interval in a trial, either the target or the interferer had components starting in sine phase, while the components of the other part were advanced in time by a fixed amount to introduce a PPA. To achieve this, the starting phase of the nth harmonic was shifted by $n\times\Delta\varphi$. In one interval the pitch pulses in the target were advanced relative to those of the interferer, and in the other interval the pitch pulses in the interferer were advanced relative to those of the target. The values of $\Delta\varphi$ were 45° (corresponding to a PPA of 1.42 ms) and 90° (corresponding to a PPA of 2.84 ms).

The complex tones were generated and bandpass filtered digitally. They were played out using a 16-bit digital-to-analog converter (CED 1401 plus), with a sampling rate which was varied between trials over the range 40 kHz $\pm$10%. As a result, the $F0$ of the digitally generated stimuli varied randomly over the range $\pm$10% between trials (this also produced a slight variation in duration and in the filter cutoff frequencies). This was done to make the stimuli as similar as possible to those used by Gockel *et al.* (2004). Stimuli were passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 17.2 kHz (slope of 96 dB/oct), and presented monaurally, using Sennheiser HD250 headphones. Subjects were seated individually in an IAC double-walled sound-attenuating booth.

C. Procedure

A two-interval two-alternative forced choice (2I-2AFC) task was used to measure percent-correct performance. The task of the subject was to identify the interval in which the pitch pulses of the target were advanced relative to those of the interferer. Correct-answer feedback was provided after every trial. The value of $\Delta\varphi$ was fixed within a block of trials. The total duration of a single session was about 2 h, including rest times. Following 1–2 h of practice, at least five (mostly 6–7) blocks of 105 trials were run for each value of $\Delta\varphi$ and each subject. The first five trials within each block were considered as ''warm-up'' trials and results from those were discarded.

D. Subjects

Six subjects participated. They ranged in age from 19 to 34 years, and their quiet thresholds at octave frequencies between 250 and 8000 Hz were within 15 dB of the ANSI (1996) standard. Stimuli were presented monaurally to the left ear for four subjects and to the right ear for the other two.

E. Results

The percent-correct scores were converted to d' values. The results were similar across subjects and Fig. 1 shows the mean results and standard errors, i.e., the error of the mean of the average score across subjects. The task did prove to be possible; the d' value was relatively small (about 0.7) for the smaller value of $\Delta\varphi$, and increased to about 1.5, indicating reasonably good discriminability, for the larger value of $\Delta\varphi$. A t-test showed that the value of d' for $\Delta\varphi=45°$ was significantly greater than zero ($t=3.3$, df$=5$, $p<0.05$). A paired-samples t-test showed that d' for $\Delta\varphi=90°$ was significantly greater than d' for $\Delta\varphi=45°$ ($t=4.85$, df$=5$, $p<0.01$).

For three out of the six subjects, we also tested whether the ability to discriminate the direction of PPA would be reduced if the starting phase of the two complexes was randomized for each presentation (maintaining a given PPA). No consistent differences were observed between performance with fixed and random starting phases.

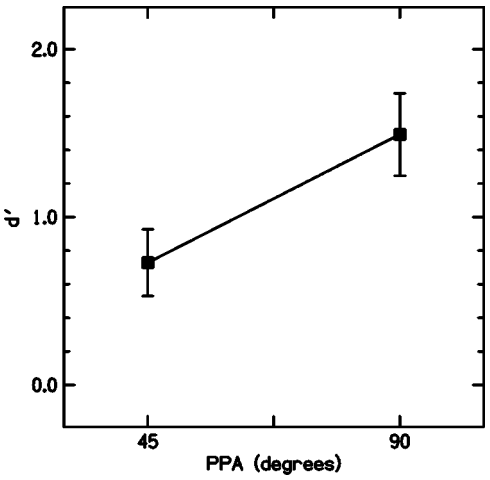


FIG. 1. Means and standard errors (across six subjects) of values of d' for discrimination of the direction of ''pitch pulse asynchrony'' (PPA) between the two parts of a complex tone with a nominal $F0$ of 88 Hz. One part, referred to as the ''target,'' was bandpass filtered between 3900 and 5400 Hz. The other part, referred to as the ''interferer'', was bandpass filtered between 1375 and 1875 Hz. The value of d' is plotted as a function of the amount of PPA, specified as $\Delta\varphi°$ (see the text for details).

In summary, the results show that, for stimuli very similar to those used by Gockel *et al.* (2004), subjects can discriminate differences in the direction of PPA across frequency regions. This makes it more plausible that, in experiment 3A of Gockel *et al.* (2004), subjects might have used PPA cues to improve $F0$ discrimination of the target in the presence of the interferer with unresolved harmonics. In the following two experiments, we assessed the effect of eliminating PPA cues, in a task very similar to that used by Gockel *et al.* (2004).

III. EXPERIMENT 2: PDI WITH RANDOM PHASE FOR TARGET

A. Stimuli

Subjects had to discriminate between the $F0$s of two sequentially presented target complex tones with a nominal $F0$ of 88 or 250 Hz and a fixed difference, $\Delta F0$, between the two tones within a trial. Each target was bandpass filtered between 3900 and 5400 Hz (3-dB down points, slopes of 48 dB/oct); thus, for both $F0$s it contained only unresolved components. The target was presented either alone (condition ''none''), or with a synchronously gated harmonic complex (the interferer) with an $F0$ which was identical to the nominal $F0$ of the target. The interferer was bandpass filtered between 1375 and 1875 Hz (3-dB down points, slopes of 48 dB/oct). For the 88-Hz $F0$, this meant that the interferer contained only unresolved harmonics, while for the 250-Hz $F0$ the interferer contained mainly resolved harmonics (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003). Apart from the component phases, other details of the stimuli were the same as for experiment 1, including the presence of the pink noise background. The $F0$ randomization was again used, in this case to discourage subjects from basing their decisions on a long-term memory representation of the sound, and to encourage them to compare the pitch of the two targets presented in each trial.

To avoid any cue provided by PPA between the target and interferer with unresolved harmonics, the target harmonics was added with random starting phase. For the target, the starting phase of each component was randomly chosen from the range 0–360° (rectangular distribution). For each target complex, 21 versions were generated in advance, each with a different random selection of starting phases, and one out of these was chosen randomly for each presentation. The components of the interferer were added in sine phase.

For all subjects, $\Delta F0$ was 18%. This was substantially higher than the 3.5% used by Gockel *et al.* (2004) in their experiment 3A. The large value of $\Delta F0$ was necessary because pilot experiments showed that $F0$ discrimination for the random-phase target was much poorer than for the sine-phase target originally used.

B. Procedure

A 2I-2AFC task was used to measure percent-correct performance. The subjects were required to indicate the interval in which the target tone had the higher $F0$. They were given practice with the target alone before taking part in runs where an interferer was present. Thus, subjects were familiar with the sound of the target. When an interferer was present, subjects were requested to focus attention on the target and to ignore the interferer as much as possible. Each interval was marked by a light, and visual feedback was provided following each response.

C. Subjects

Six subjects participated, one of whom was the first author. They ranged in age from 22 to 42 years, and their quiet thresholds at octave frequencies between 250 and 8000 Hz were within 15 dB of the 1996 ANSI standard. All of the subjects had previous experience in pitch discrimination experiments. Stimuli were presented monaurally to the left ear for three subjects and to the right ear for the other three.

D. Design

The total duration of a single session was about 2 h, including rest times. At least five (mostly 6–7) blocks of 105 trials were run for each condition and subject. Within a block of 105 trials, the condition was kept constant. The first five trials within each block were considered as warm-up trials and results from those were discarded. The order of the conditions was counterbalanced within and across subjects. One block was run for each condition in turn, before additional blocks were run in any other condition.

E. Results and discussion

The percent-correct scores were converted to d' values. The results were similar across subjects and Fig. 2 shows the mean results and standard errors. The left two columns show results for the target with a nominal $F0$ of 88 Hz (interferer with unresolved harmonics), and the right two columns show results for the target with a nominal $F0$ of 250 Hz (interferer

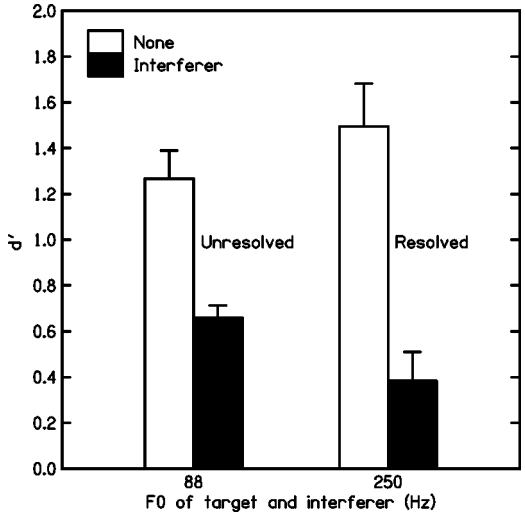


FIG. 2. Means and standard errors (across six subjects) of values of d' for $F0$ discrimination of targets filtered between 3900 and 5400 Hz presented either alone (open bars) or with an interferer (filled bars). The harmonics of the target were added in random phase. The harmonics of the interferer were added in sine phase. $\Delta F0$ for the target was 18% for all subjects. The $F0$ of the interferer was equal to the nominal $F0$ of the target, which was either 88 Hz (left pair of bars) or 250 Hz (right pair of bars).

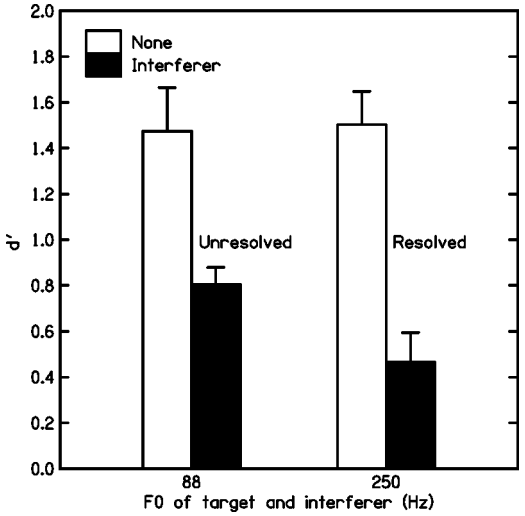


FIG. 3. As Fig. 2, except that the harmonics of the interferer were added in random phase.

with resolved harmonics). The white and black columns show d' in the absence and in the presence of an interferer, respectively. A PDI effect occurred for both interferers. The important finding here is that PDI was greater when the interferer contained resolved harmonics than when it contained unresolved harmonics, despite the elimination of PPA cues for the interferer with unresolved harmonics. A repeated measures two-way analysis of variance (ANOVA) was conducted with factors $F0$ and presence or absence of an interferer. The ANOVA showed a significant main effect of the presence of the interferer [$F(1,5)=29.7$, $p=0.003$] and a significant interaction between $F0$ and presence of the interferer [$F(1,5)=11.1$, $p=0.021$]. There was no main effect of $F0$.

In summary, the results show that an interferer containing resolved harmonics produced more PDI than an interferer containing unresolved harmonics, even when cues relating to PPA were not available.

IV. EXPERIMENT 3: PDI WITH RANDOM PHASE FOR BOTH TARGET AND INTERFERER

A. Rationale

In experiment 2, the harmonics of the interferer were added in sine phase and the harmonics of the target were added in random phase. While this would have eliminated cues related to PPA, it is possible that, for the interferer with unresolved harmonics, the difference in peak factor between the interferer and the target led to a difference in timbre which could be used as a cue for perceptual segregation of the interferer and target. It has been shown that differences in phase spectra of this type can be used for perceptual segregation of successive tones (Roberts *et al.*, 2002; Stainsby *et al.*, 2004), but it is not clear whether they can be used for segregation of simultaneous tones in different frequency regions. In experiment 3, possible cues related to the difference in peak factor between the target and the interferer were eliminated by adding the harmonics in random phase for both the interferer and the target.

B. Stimuli, procedure, and subjects

The stimuli were the same as in experiment 2, except that the harmonics in the interferer also had random starting phases. The procedure and subjects were the same as for experiment 2.

C. Results and discussion

The percent-correct scores were converted to d' values. The results were similar across subjects; Fig. 3 shows the mean results and standard errors. As in experiment 2, the PDI effect was greater when the interferer contained resolved harmonics than when it contained unresolved harmonics. A repeated measures two-way ANOVA (with factors $F0$ and presence or absence of an interferer) showed a significant main effect of the presence of the interferer [$F(1,5)=24.1$, $p=0.004$] and a significant interaction between $F0$ and presence of the interferer [$F(1,5)=11.1$, $p=0.012$]. There was no main effect of $F0$.

In summary, the results show that an interferer containing resolved harmonics produced more PDI than an interferer containing unresolved harmonics, even when cues relating to PPA, or to differences in peak factor between the target and the interferer, were not available.

V. GENERAL DISCUSSION

The results of experiment 1 showed that subjects could discriminate differences in the direction of PPA across frequency regions. This made it more plausible that, in experiment 3A of Gockel *et al.* (2004), subjects might have used PPA cues to improve $F0$ discrimination of the target in the presence of the interferer with unresolved harmonics.

The results of experiments 2 and 3 showed that PDI was greater when the interferer contained resolved harmonics than when it contained unresolved harmonics, even when

cues related to PPA for the latter were eliminated. However, the magnitude of the PDI effect produced by the interferers with unresolved harmonics in experiments 2 and 3 was somewhat larger than found by Gockel *et al.* (2004) in their experiment 3A. This comparison, and the finding of our experiment 1 that subjects are sensitive to the direction of PPA differences across frequency regions, are consistent with the idea that, in experiment 3A of Gockel *et al.*, PPA cues may have somewhat reduced the PDI produced by the interferer with unresolved harmonics. The reduction may have been small because the PPA cues were relatively weak. Nevertheless, our results suggest that the smaller PDI produced by an interferer containing unresolved harmonics is not caused solely by the use of PPA cues, or by differences in the peak factor of the target and interferer. These results provide further evidence against the idea that there are two pitch mechanisms, one for resolved harmonics and one for unresolved harmonics, whose outputs can be independently accessed.

The most plausible explanations for the greater PDI produced by the interferer with resolved harmonics in our experiments 2 and 3 are (1) The resolved interferer has a greater pitch salience, and the amount of PDI is determined at least partly by the pitch salience of the interferer; (2) There is a special status of resolved harmonics with regard to pitch processing, since those harmonics usually dominate the pitch percept (Plomp, 1967; Ritsma, 1967; Moore *et al.*, 1985). A third, less likely possibility is that the magnitude of PDI increases with increasing $F0$, independent of whether the interferer is resolved or unresolved. Even though there is no obvious reason why this should be the case, a preliminary experiment was conducted to address this possibility. Target tones with nominal $F0$s of 88 and 250 Hz were bandpass filtered between 3900 and 5400 Hz. Interferers with $F0$s of 88 and 250 Hz were bandpass filtered between 125 and 625 Hz so that both contained resolved components. There was no consistent trend for PDI to be larger for the 250-Hz $F0$ than for the 88-Hz $F0$ condition.

With regard to explanation (1), it is noteworthy that the amount of PDI produced by the interferer with unresolved harmonics was almost unaffected by whether the harmonics of the interferer were added in sine phase (experiment 2) or in random phase (experiment 3). The salience of the pitch of the interferer would have been lower when the harmonics was added in random phase than when they were added in sine phase (Bilsen and Ritsma, 1970; Warren and Bashford, 1981; Roberts *et al.*, 2002), but this hardly affected the amount of PDI. At first sight, this appears to suggest that the pitch salience of the interferer does not determine the amount of PDI. However, it is possible that the difference in phase spectrum of the target and the (unresolved) interferer that was present in experiment 2 led to a difference in timbre (Bilsen and Ritsma, 1970; Warren and Bashford, 1981; Roberts *et al.*, 2002), and that this in turn provided a cue for perceptual segregation of the target and interferer. Since PDI is known to be affected by perceptual grouping factors (Gockel *et al.*, 2004), perceptual segregation produced by the timbre difference between the target and interferer might have offset the effect of the greater pitch salience of the interferer in experiment 2, leading to similar PDI in experiments 2 and 3. Therefore, we can conclude only that the pitch salience of the interferer is not the only factor influencing the magnitude of PDI.

VI. SUMMARY AND CONCLUSIONS

Gockel *et al.* (2004) showed previously that PDI was greater when the interferer contained resolved harmonics than when it contained only unresolved harmonics. In their experiment, both target and interferer had components added in sine phase, so it was possible that the lower PDI produced by the interferer with unresolved harmonics was caused by the use of cues arising from PPA between the target and interferer. The first experiment reported here assessed whether subjects are sensitive to the direction of PPA across frequency regions. The results showed that this was the case. To determine whether the use of PPA cues was responsible for the smaller PDI produced by an interferer with unresolved harmonics, the effect of reducing PPA cues was assessed by randomizing the starting phases of either the target only (experiment 2) or both the target and interferer (experiment 3). For both experiments, the interferer containing resolved harmonics produced more PDI than the interferer containing unresolved harmonics. These results show that, even when PPA cues are not available, an interferer with unresolved harmonics produces less PDI than an interferer with resolved harmonics. The magnitude of PDI may be related to the pitch salience of the interferer, but this cannot be the only factor involved, since the amount of PDI was similar for experiments 2 and 3, but the interferer with unresolved harmonics would have had a more salient pitch in experiment 2 than in experiment 3.

ACKNOWLEDGMENTS

This work was supported by EPSRC Grant GR/R65794/01 and by the MRC. We thank two anonymous reviewers and John Grose for helpful comments on a previous version of this manuscript.

Alcántara, J. I., Moore, B. C. J., Glasberg, B. R., Wilkinson, A. J. K., and Jorasz, U. (**2003**). "Phase effects in masking: Within- versus across-channel processes," J. Acoust. Soc. Am. **114**, 2158–2166.

ANSI (**1996**). ANSI S3.6-1996, "American National Standard Specification for Audiometers" (American National Standards Institute, New York).

Assmann, P. F., and Summerfield, A. Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.

Bilsen, F. A., and Ritsma, R. J. (**1970**). "Some parameters influencing the perceptibility of pitch," J. Acoust. Soc. Am. **47**, 469–475.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge, MA).

Carlyon, R. P. (**1994**). "Detecting pitch-pulse asynchronies and differences in fundamental frequency," J. Acoust. Soc. Am. **95**, 968–979.

Carlyon, R. P. (**1998**). "Comments on "A unitary model of pitch perception" [J. Acoust. Soc. Am. **102**, 1811–1820 (1997)]," J. Acoust. Soc. Am. **104**, 1118–1121.

Carlyon, R. P., and Shackleton, T. M. (**1994**). ''Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?,'' J. Acoust. Soc. Am. **95**, 3541–3554.
Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
de Cheveigné, A. (**1993**). ''Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing,'' J. Acoust. Soc. Am. **93**, 3271–3290.
Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). ''Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms,'' J. Acoust. Soc. Am. **116**, 1092–1104.
Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.
Grimault, N., Micheyl, C., Carlyon, R. P., and Collet, L. (**2002**). ''Evidence for two pitch encoding mechanisms using a selective auditory training paradigm,'' Percept. Psychophys. **64**, 189–197.
Houtsma, A. J. M., and Smurzynski, J. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.
Meddis, R., and Hewitt, M. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and Hewitt, M. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Micheyl, C., and Oxenham, A. J. (**2004**). ''Sequential *F*0 comparisons between resolved and unresolved harmonics: No evidence for translation noise between two pitch mechanisms,'' J. Acoust. Soc. Am. **116**, 3038–3050.
Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).
Moore, B. C. J. (**2002**). ''Interference effects and phase sensitivity in hearing,'' Philos. Trans. R. Soc. London, Ser. A **360**, 833–858.
Moore, B. C. J., and Glasberg, B. R. (**1988**). ''Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral and bilateral cochlear impairments,'' in *Basic Issues in Hearing*, edited by H. Duifhuis, H. Wit, and J. Horst (Academic, London).
Moore, B. C. J., and Ohgushi, K. (**1993**). ''Audibility of partials in inharmonic complex tones,'' J. Acoust. Soc. Am. **93**, 452–461.
Moore, B. C. J., and Peters, R. W. (**1992**). ''Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity,'' J. Acoust. Soc. Am. **91**, 2881–2893.
Moore, B. C. J., and Rosen, S. M. (**1979**). ''Tune recognition with reduced pitch and interval information,'' Q. J. Exp. Psychol. **31**, 229–240.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.
Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). ''Frequency and intensity difference limens for harmonics within complex tones,'' J. Acoust. Soc. Am. **75**, 550–561.
Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.
Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.
Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.
Plomp, R., and Mimpen, A. M. (**1968**). ''The ear as a frequency analyzer. II,'' J. Acoust. Soc. Am. **43**, 764–767.
Rhode, W. S., and Recio, A. (**2001**). ''Basilar-membrane response to multicomponent stimuli in chinchilla,'' J. Acoust. Soc. Am. **110**, 981–994.
Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). ''Primitive stream segregation of tone sequences without differences in *F*0 or passband,'' J. Acoust. Soc. Am. **112**, 2074–2085.
Scheffers, M. T. M. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' Ph.D. thesis, Groningen University, The Netherlands.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Stainsby, T. H., Moore, B. C. J., and Glasberg, B. R. (**2004**). ''Auditory streaming based on temporal structure in hearing-impaired listeners,'' Hear. Res. **192**, 119–130.
Summerfield, A. Q., and Assmann, P. (**1987**). ''Modelling the perception of concurrent sounds,'' Meeting of Exp. Psychol. Soc. **1987**.
Summerfield, A. Q., and Assmann, P. F. (**1991**). ''Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony,'' J. Acoust. Soc. Am. **89**, 1364–1377.
Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.
Warren, R. M., and Bashford, Jr., J. A. (**1981**). ''Perception of acoustic iterance: Pitch and infrapitch,'' Percept. Psychophys. **29**, 395–402.

Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English

Taehong Cho[a)]
Division of English Language and Literature, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul, 133-791, Korea, and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 20 September 2002; revised 5 January 2005; accepted 5 January 2005)

In this study the effects of accent and prosodic boundaries on the production of English vowels (/ɑ,i/), by concurrently examining acoustic vowel formants and articulatory maxima of the tongue, jaw, and lips obtained with EMA (Electromagnetic Articulography) are investigated. The results demonstrate that prosodic strengthening (due to accent and/or prosodic boundaries) has differential effects depending on the source of prominence (in accented syllables versus at edges of prosodic domains; domain initially versus domain finally). The results are interpreted in terms of how the prosodic strengthening is related to phonetic realization of vowel features. For example, when accented, /i/ was fronter in both acoustic and articulatory vowel spaces (enhancing [−back]), accompanied by an increase in both lip and jaw openings (enhancing sonority). By contrast, at edges of prosodic domains (especially domain-finally), /i/ was not necessarily fronter, but higher (enhancing [+high]), accompanied by an increase only in the lip (not jaw) opening. This suggests that the two aspects of prosodic structure (accent versus boundary) are differentiated by distinct phonetic patterns. Further, it implies that prosodic strengthening, though manifested in fine-grained phonetic details, is not simply a low-level phonetic event but a complex linguistic phenomenon, closely linked to the enhancement of phonological features and positional strength that may license phonological contrasts. [DOI: 10.1121/1.1861893]

PACS numbers: 43.70.Aj, 43.70.Eq [AL] Pages: 3867–3878

I. INTRODUCTION

One of the fundamental presuppositions shared among researchers in the fields of prosodic phonology and the phonetics–prosody interface is that speech utterances are produced in a ''hierarchically organized structure of phonologically defined constituents and heads'' (Beckman, 1996) with higher constituents being decomposed into lower constituents. The prosodic structure has been considered as an essential element in speech production, as it is not only a source of subphonemic variation, but also it constrains the phonological shape of a segment by forming different domains of application of phonological rules (cf. Selkirk, 1986; Jun, 1998). The present study explores how prosodic structure influences the production of the English vowels /ɑ,i/, and how prosodic strengthening (i.e., temporal and/or spatial expansion of articulation due to accent and/or prosodic boundaries) can be related to the enhancement of phonological vowel features such as sonority and place features, which may ultimately shed light on the relationship between prosodic strengthening and linguistic contrast.

There are various approaches to determining prosodic groupings [see Shattuck-Hufnagel and Turk (1996) and Jun (1998) for a general review and a discussion]. One line of research assumes that the prosodic structure of an utterance is a grammatical entity in its own right, and it is realized on the surface by distinctive phonetic patterns, primarily via suprasegmental features such as pitch and temporal structure (Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988; Jun, 1993, and Beckman, 1996). The model of prosodic structure in this approach generally assumes that lower domains (e.g., Prosodic Words) are grouped into immediately higher levels [e.g., the Intermediate Phrases (ip)], eventually forming the Intonational Phrase (IP), the highest phonological constituent.

In seeking the phonetic correlates of prosodic structure, researchers have traditionally paid special attention to the end of prosodic domains. For example, prosodic units larger than a (phonological) word are generally defined in terms of the distribution of phrase-final intonational contrasts (e.g., Bolinger, 1970; Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988) and final lengthening (e.g., Edwards *et al.*, 1991; Beckman *et al.*, 1992; Wightman *et al.*, 1992). As for other phonetic events at the end of prosodic domains, it has been suggested that domain-final vowels may be produced with greater articulatory magnitude, resulting in greater V-to-C displacement (Fougeron and Keating, 1997; Cho, 2002; in-press; Tabain, 2003). Moreover, Cho (2004) has shown that boundary strength is reflected in cumulative V-to-V coarticulatory resistance: domain-final vowels resist coarticulation with the following vowels to a greater degree across a higher prosodic boundary than across a lower one. These studies suggest that lengthening is just one of many phonetic phenomena occurring at the ends of phrases, which, taken together, may be thought of as a type of prosodic strengthening to mark the ends of prosodic domains. [See Byrd and Saltzman (2003) for simulations in the framework of the mass-spring gestural model, which demonstrated that a clock-slowing implementation may induce both temporal and spatial expansion at a stronger prosodic boundary.]

[a)]Electronic mail: tcho@hanyang.ac.kr

Other recent studies, however, have shown that prosodic domains may also be signaled by extreme articulation at the left edges of the domains. For instance, consonants at the beginnings of phrases are more constricted than consonants in the middle of phrases. This phenomenon, known as *domain-initial strengthening*, has been documented in a series of electropalatographic studies in English, French, Korean, and Taiwanese (Fougeron and Keating, 1997; Cho and Keating, 2001; Fougeron, 2001; see also Keating *et al.* (1999) for a cross-linguistic comparison). (See also Cho and Jun, 2000 and Cho and McQueen, in press, for differential phonetic effects of domain-initial strengthening in connection with language-specific phonetic feature systems, in Korean and Dutch, respectively.) Domain-initial strengthening on a word level has also been observed acoustically by Turk and Shattuck-Hufnagel (2000).

Finally, prosodic strengthening effects can also be found in accented syllables, which may be thought of as the heads of Intermediate Phrases (see Beckman and Edwards, 1994). In addition to receiving distinctive tonal properties (such as pitch accents), segments in accented syllables are also more strongly articulated (e.g., Beckman *et al.*, 1992; de Jong, 1995a, b; Erickson, 2002; Cho, in press).

These three strong prosodic positions (i.e., domain-final, domain-initial, and accented positions), all of which are subject to prosodic strengthening, are the loci of the present investigation: it examines the phonetic realization of the vowels /ɑ,i/ both acoustically (as reflected in the F1−F2 vowel space) and articulatorily (as reflected in the maxima of the tongue, the jaw, and the lip opening). (Note that the jaw and lip opening maxima are likely to represent the degree of the vocal track opening that is related to sonority expansion, whereas the tongue maxima together with the acoustic data can be interpreted primarily as an index of how place features are phonetically realized).

While segments in prosodically strong positions are generally taken to be characterized by distinct phonetic patterns, most previous studies discussed above have focused on just one articulator or one or two prosodic positions, leaving it not entirely clear whether the phonetic signatures of prosodic structure differ in kind depending on the prosodic position. This leads to questions regarding whether and how articulation under accent differs from articulation at the edges of prosodic domains, and whether and how domain-initial articulation differs from domain-final articulation.

The second question is related to prosodic strengthening and linguistic contrast. Although prosodically strong positions have been regarded as ''prominent'' or ''privileged'' positions in which sonority is expanded (e.g., Beckman *et al.*, 1992) and phonemic contrasts are likely to be maximized (e.g., de Jong, 1995a, b) or maintained (e.g., Beckman, 1998; Barnes, 2002; cf. Steriade, 1999), our knowledge on the relationship between the actual phonetic consequences of prosodic strengthening and linguistic contrast has been extremely limited. The present study therefore sets out to determine how prosodic strengthening is related to linguistic contrast. Specifically, it investigates how prosodic strengthening associated with different prosodic positions is related to the enhancement of phonological features of the vowels /ɑ,i/, which could result in linguistic contrast being maximized, but possibly differentially depending on the source of prominence (e.g., in accented syllables versus at edges of prosodic domains; domain initially versus domain finally). Below are some specific hypotheses to be assessed in connection with the two main research questions outlined above.

With respect to accent-induced prominence, Edwards and Beckman (1988) and Beckman *et al.* (1992) proposed that an accent has the effect of enhancing a segment's intrinsic sonority (the Sonority Expansion Hypothesis). Under this account, the influence of accent on vowels is to make them more vowel-like by opening the vocal tract more; the influence on consonants is to make them more consonant-like by closing the vocal tract more tightly. Subsequently, de Jong (1995a) proposed that stress leads to hyperarticulation, enhancing the distinctive features of segments in a way that may maximize lexical distinctions, and sonority is just one of many distinctive features that are enhanced. In support of this localized hyperarticulation hypothesis, he showed that the tongue position for an English vowel /ʊ/ is backer, being accompanied by a lower jaw position,[1] which indicates both sonority expansion (jaw lowering) and the enhancement of place features ([+back] by the tongue backing).

Under this hypothesis, it is of particular interest to examine /ɑ/ and /i/ at the same time because tongue lowering for /ɑ/ could be interpreted as suggesting an enhancement of both nonsonority (place) feature (e.g., [+low]) and sonority feature, whereas tongue raising for /i/ could be interpreted as an enhancement of the place feature [+high], but in conflict with enhancement of the sonority feature as it reduces the size of the vocal track opening. In the latter case, however, sonority expansion may still be attained by making the jaw and the lip openings larger independently of the tongue raising. Thus, under the hyperarticulation hypothesis, it is predicted that the tongue is involved primarily in the enhancement of the nonsonority (place) features, whereas the jaw and the lips are involved primarily to achieve sonority feature enhancement, independently of the tongue.

With respect to boundary-induced prominence, Farnetani and Vayra (1996) hypothesize that while prominence under accent fits the localized hyperarticulation hypothesis, prominence at edges fits sonority expansion, which would heighten ♯CV or V♯C contrast (where ♯=some prosodic boundary). Thus, one can predict that prosodic strengthening at the edges of prosodic domains is different in kind from that associated with an accent, such that the former is driven primarily by the sonority expansion, but not by the enhancement of place features. If this were the case, all three articulators would converge on sonority expansion to be reflected in increased lip/jaw opening and tongue lowering for both vowels /ɑ/ or /i/ at a higher prosodic boundary. It should be noted, however, that place features may still be opportunistically enhanced to the extent that is not in conflict with sonority expansion. The enhancement of the feature [+high] for /i/ (as may be reflected in tongue raising together with decreased lip and jaw openings) would be suppressed, as it would result in affecting the sonority expansion antagonistically. On the other hand, the enhancement of the place feature [+low] for /ɑ/ goes hand in hand with sonority expan-

sion: the articulatory maneuvers for sonority expansion of /ɑ/ (e.g., increased lip and jaw openings and the tongue lowering) would effectively result in the enhancement of [+low].

This hypothesis will be tested separately for domain-initial and domain-final vowels since there have been different predictions made in the literature. For example, Fougeron and Keating (1997) speculate that strengthening may be very local to the domain edge such that the boundary-induced strengthening effect may be evident primarily in the domain-final vowel and the domain-initial consonant, resulting in a greater V-to-C displacement across a larger prosodic boundary. Since a domain-initial vowel in ♯CV is not strictly domain-initial, the sonority expansion effect might be smaller, or even absent, for domain-initial vowels.

In a traditional phonological framework (e.g., SPE, Chomsky and Halle 1968), vowels have often been defined in terms of the height and backness of the tongue (e.g., [+high, -back] for /i/ and [+low, +back] for /ɑ/). However, some researchers have proposed that vowel features may be better understood in acoustic/auditory terms, for example in a F1-F2 plane (e.g., Neary, 1977, Flemming, 1995, Ladefoged, 2000). However, it should be noted here that in some other frameworks, vowels are described in terms of constriction degree and location (Wood, 1979, 1982; Jackson, 1988; cf. Browman and Goldstein, 1992). In this framework, /i/ is characterized as having a narrow constriction in the region of the hard palate and /ɑ/ as having a narrow constriction in the lower pharynx. As described below, the articulatory component of the present study has been designed to identify the tongue position in a maxilla-based coordinate system defined by the occlusal plane without making specific reference to the opposing walls. Thus, it will be difficult to evaluate the prosodic strengthening effect on articulation in terms of the constriction degree and location. Similarly, the articulatory data in the present study provide positions of pellets attached on the tongue dorsum area, but not the tongue root area, which allows for only indirect assessment of the tongue's backness for /ɑ/. Thus, the articulatory data with respect to the tongue's physical backness in this study must be interpreted as being indicative rather than conclusive.

II. EXPERIMENT

A. Speech material and speakers

Each sentence in the corpus included two test syllables (domain-final/domain-initial), yielding a $C_1V_1\sharp C_2V_2$ sequence (♯=some prosodic boundary) across two English words, as seen in sample sentences given in Table I. C_1 and C_2 were always /b/, whose articulation is known to minimally interfere with the vocalic lingual articulation, and V_1 and V_2 were homorganic (/i♯bi/ and /ɑ♯bɑ/). As for prosodic variables, the boundary between the test syllables was varied from the Intonational Phrase boundary (IP), to the Intermediate Phrase boundary (ip), to the Word boundary (Wd). Accentuation was also manipulated in preboundary and postboundary syllables, resulting in four pairs: ACC♯ACC, ACC♯UNACC, UNACC♯ACC, UNACC♯UNACC. Such a manipulation of prosodic factors yields three prosodic variables: (a) strength of prosodic boundary; (b) accentuation of syllables adjacent to the boundary (accented, unaccented); (c) position-in-domain of test syllables (initial, final). This yielded a total of 48 different sequences (3 prosodic boundaries×2 accentual patterns×2 accentual patterns in the neighboring vowel×2 edges×2 vowel types).

TABLE I. Sample sentences containing the sequence /bɑ♯bɑ/ with different prosodic boundaries (IP,ip,Wd) and accentual patterns. The words receiving accent are highlighted in bold.

♯=Word boundary:		
(a) ACC.-UNACC.		
Prompt:	Did you just say "Little **Boo** bopped the girl last night"?	
Target:	No, "Little **Bah** ♯ bopped the girl"	
rendition:	(L+)H*	L−L%
♯=Intermediate or Intonational Phrase boundaries (ip or IP):		
(b) ACC.-UNACC.		
Prompt:	Did you say "Little **Boo** bopped the **boy** last night"?	
Target:	No, "Little **Bah** ♯ bopped the **girl**."	
rendition 1:	(L+)H*L−	(L+)H* L−L%
rendition 2:	(L+)H*L−L%	(L+)H* L−L%

Six native speakers of American English participated in the experiment, but only five were included in this study due to technical errors (see below). Participants were either linguistics students or postdoctoral fellows at UCLA. In order to control for the variation in rounding in the low vowel, speakers whose dialect lacked the phoneme /ɔ/ were chosen, so that the vowel in the test word *bop* was produced as unrounded low back vowel /ɑ/. Speakers were all trained in the production of English sentences in the ToBI (Tone and Break Index) framework (see Silverman *et al.*, 1992; Beckman and Ayers, 1997) prior to the experiment. Before the actual recording date, each speaker participated in an approximately two-hour long practice session.

B. Procedures

Accent patterns were recorded in lexically contrasting discourse frames, as in Table I, in which the two word target sequence is /bɑ♯bɑ/ in *Little Bah bopped the girl*. (/bi♯bi/ sequence tokens were produced in similar discourse frames as in *Donna B. beeped at him*.) The words highlighted in bold in the table received the accent. The prompt was read silently by the speaker to cue the intended accent patterns, which were provided using partial ToBI transcriptions in the script (see below). For a balanced number of tokens for ip and IP, speakers were asked to produce two different renditions for each sentence designed for inducing phrasal boundaries as shown in (b) in Table I: one with an ip boundary (*rendition 1*) and one with and IP boundary (*rendition 2*). Speakers had no difficulty producing two versions of each sentence (ip and IP), given the practice session and previous phonetic training. Each sentence was read twice in succession at their comfortable rate of speech and the entire list was read twice, for a total of four repetitions per sentence. This yielded a total of 960 sentence tokens for analysis (48 sentence types×5 speakers×4 repetitions).

An EMA system (Carstens Articulograph AG 100) was used with seven transducer coils to track articulatory movements (cf. Hoole, 1996). Two reference transducers were placed on the nose and upper gumline, or maxillary incisor, both of which were used by the EMA analyzing software to

establish a maxilla-based coordinate system and to correct for head movement inside the helmet. The remaining five transducers were located on articulators: Two transducer coils were placed on the tongue: one at the tongue dorsum (TD) about 5 cm from the tongue tip and one at the tongue midsection (TM), about 2.5 cm from the tongue tip.

Next, in order to align the articulatory space to the maxillary occlusal plane, a flat plastic bite plate (approximately the size of bank card, custom-fit to the size of the subject's mouth) was used with two transducer coils attached along the midline that corresponds to the midsagittal line. The articulatory space was rotated so that the *x* axis was the maxillary occlusal plane. Based on such a rotation, a maxilla-based coordinate system was established in which the *x* axis is parallel to the maxillary occlusal plane with the *y* axis being perpendicular to that at the junction of the occlusal plane and the central maxillary incisor, and this is consistent across speakers. [A similar data processing procedure was employed by Tabain (2003); see Westbury (1994) for a further discussion about the usefulness of the occlusal plane.] The EMA data were sampled at 500 Hz. The obtained kinematic signals were then submitted to low-pass filtering with a filter cutoff of 50 Hz, using Tailor (Carstens' data processing program; see http://linguistics.ucla.edu/faciliti/facilities/physiology/ema.htm.

The relevant $C_1V_1\#C_2V_2$ portion of the audio recording was transcribed, with the aid of an acoustic display, by two trained ToBI transcribers (one the author) following the criteria set forth in the ToBI transcription system. In general, accents receive either H* or L+H*, and three prosodic boundaries were identified: the IP boundary (marked by a boundary tone and a break index 4); the ip boundary (marked by a break index 3, a phrasal tone, and no boundary tone); the Word boundary (marked by a break index 1, in the middle of an Intermediate Phrase). The two transcribers identified identical locations of accent in every token of the entire dataset. The only difference between the two transcribers came from a choice between the Intonational Phrase boundary and the Intermediate Phrase boundary, and between L+H* and H* for accented words. Because the difference between IP and ip boundaries is an important experimental variable in this study, only tokens whose renditions were agreed on by the two transcribers (94.3%) were used for analysis.[2]

The extreme points of the tongue mid (TM) and tongue dorsum (TD) transducers were primarily identified from minima in the tangential velocity signal, following Löfqvist, Gracco, and Nye (1993) and Löfqvist (1999). These points were cross-checked by inspecting a sagittal display of the tongue movement trajectories, in which ''turn-around'' points could be found. In the usual case, these correspond to the tangential velocity minima. However, cross-checking was especially useful when there was more than one tangential velocity minimum or when there was a temporal offset between the *x* and *y* maxima. In such cases, as suggested by a reviewer, the maxima in the *x* and *y* dimensions were taken separately at different times to record the actual extreme *x* and *y* position values during vowel articulation. Out of 960 tokens, 247 tokens (25.7%) showed $x-y$ timing misalignment, leading to a mean ± 8.9 ms temporal offset for the entire tokens.

The measurements for the tongue midsection and dorsum transducers were always taken at the same point in time. In most cases, the tangential velocity minima for the two transducers were synchronized. But, for the sake of consistency, the dorsum maxima were taken in synchrony with the midsection maxima. The extraction of the extreme points for the jaw during vowels was performed by taking the maximum Euclidean distance between the transducer at the lower gumline and the transducer at the upper gumline (reference point). Similarly, the extreme points for the lips during vowels were taken from the lip aperture profile as the maximum Euclidean distance between the upper and lower lip transducers. The jaw and lip opening measurements were aligned in time. Finally, the formant values were obtained from the acoustic signals sampled at 16 000 Hz in synchrony with tongue maximum points. F1 and F2 were determined from the LPC spectra (with 18 or 20 coefficients) superimposed on FFT spectra with a 512-point (32 ms) frame window centered around the tongue maximum points.

Statistical evaluation of the systematic influence of prosodic factors was primarily based on repeated measures (RM) Analyses of Variance. The within-subject factors considered were Prosodic Boundary (IP,ip,Wd) and Accent (ACC, UNACC). The factor Speaker was added as a between-subject factor to the analysis in order to take into account the speakers' contribution to any observed significance effect. Thus, a significant interaction between a prosodic factor and the factor Speaker would imply speaker-by-speaker differences. Therefore, whenever there is such a significant interaction, remarks on each speaker behavior will be made, in comparison with the overall pattern across speakers. However, adding the between-subject factor Speaker requires that each speaker contributes multiple scores per condition, making it impractical to average the data over repetitions. To compensate for this, the alpha level for significance was set more conservatively at 1% ($p<0.01$). Finally, to avoid violating the sphericity assumption, Huynh–Feldt corrected degrees of freedom (thus, often fractional values) were used in generating F ratio and *p* values.

III. RESULTS

A. Domain-final vowels ($V_1\#$)

1. Effect of accent on V_1

a. Accent effect on V_1 *F1 and F2 and tongue maxima*

The results are summarized in Fig. 1. Before delving into the detailed results, a few notes regarding the figure formats are in order. Each value in the figure (and other figures below) is the mean of speaker means with data pooled across other conditions (in this case, Boundary Type and Accent of the adjacent vowel). The F1–F2 formant plots shown in Fig. 1(a) and elsewhere have been drawn with the UCLA PlotFormants software. The scales are in accordance with the Bark scale, and the ellipses are drawn with radii of two standard deviations of the raw data along the axes of the first two principal components.

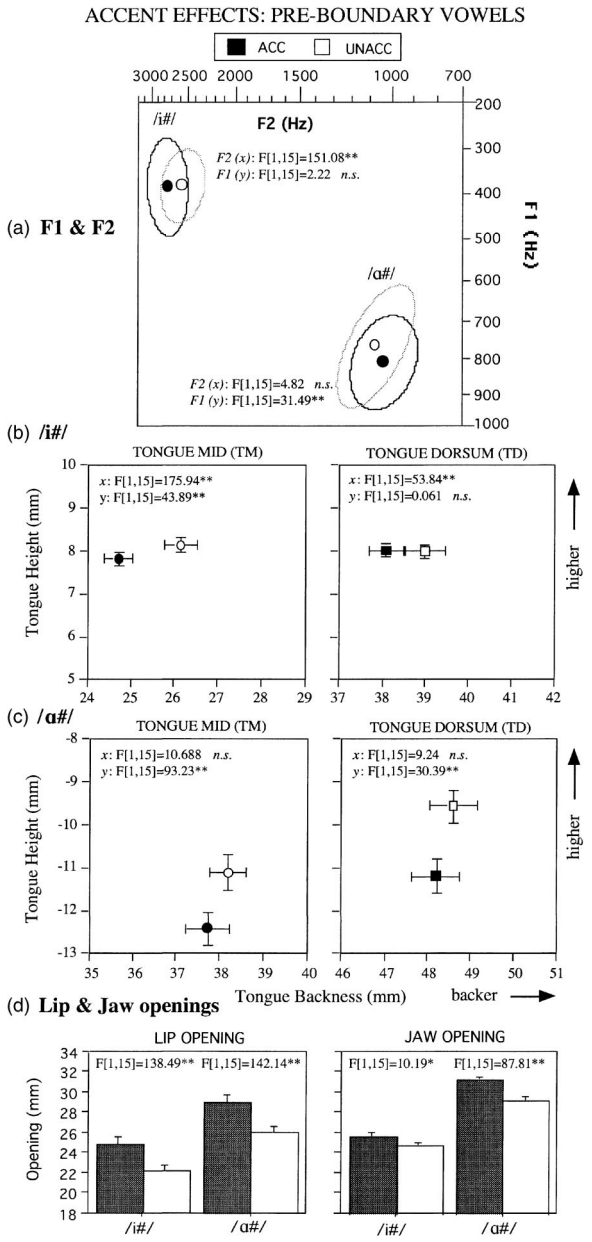


FIG. 1. Effect of Accent on F1 and F2 (a), the tongue maxima /i♯/ and /ɑ♯/ (b), (c), and the lip and jaw opening maxima (d) in the domain-final position. (* refers to $p<0.01$;** refers to $p<0.001$; $N=30$.) Note that each value in the figure (and other figures below) is the mean of speaker means with data pooled across other conditions (in this case, across prosodic boundary conditions and accent conditions of the adjacent domain-initial vowel). Thus, N is 30 (5 speakers×3 prosodic boundaries×2 accent conditions of the adjacent vowel).

Now consider /i♯/. Results of RM ANOVAs showed that /i♯/ was significantly fronter when accented versus unaccented in both the acoustic (F1–F2) and the articulatory vowel spaces [Figs. 1(a), 1(b)]. There were no interactions with Speaker (showing consistency across speakers).

In the vertical dimension, however, there was no main accent effect on /i♯/ in either the acoustic or the articulatory vowel space [Figs. 1(a), 1(b)]. However, there were significant interactions with Speaker ($F[8,30]=32.12$, $p<0.001$ for F1; $F[8,30]=8.21$, $p<0.01$ for TM; $F[8,30]=6.39$, $p<0.01$ for TD). Two speakers showed lowering effects [S1, in both the acoustic (F1) and the articulatory (TM) dimensions; S2, only in the articulatory (TD) dimension]. On the other hand, two other speakers (S4,S5) showed the opposite pattern (i.e., raising) in the acoustic (F1) dimension, but not in the corresponding articulatory dimension.

For /ɑ♯/, there were reliable acoustic and articulatory lowering effects when accented [Figs. 1(a), 1(c)] with no interactions with the Speaker. Contrary to the robust lowering effect, however, in the horizontal dimension /ɑ♯/ showed no consistent backing effect [Figs. 1(a), 1(c)]. The backing effect was found in two speakers (S1,S3) only in the acoustic dimension (F2), while no speakers showed corresponding backing in the articulatory dimension.

In summary, the high/front /i♯/, when accented, is consistently fronter in both acoustic and articulatory vowel spaces but not necessarily higher: Only two speakers (S1,S2) showed accent-induced raising in both acoustic and articulatory dimensions and two others (S4,S5) showed the opposite in the acoustic dimension. The low/back vowel /ɑ♯/ is consistently lower when accented, but backer in two speakers (S1,S3) only in the acoustic dimension (F2). (See Sec. IV A for a discussion on this).

b. Accent effect on V*1 lip and jaw opening maxima* Both lip and jaw opening maxima were reliably larger when accented than unaccented regardless of the vowel type [Fig. 1(d)]. There were no interactions with the Speaker, showing a consistent pattern across speakers.

2. Effect of boundary type on V_1

a. Boundary effect on F1 and F2 and tongue maxima in V_1 For /i♯/, there were significant accent-induced raising effects [Figs. 2(a), 2(b)]. /i♯/ was higher before a higher boundary in both acoustic (F1) and articulatory (tongue height) dimensions (IP=ip>Wd, $p<0.01$). There were no significant interactions with the Speaker. [All five speakers showed either a two-way (IP>Wd or ip>Wd or IP>ip) or a three-way (IP>ip>Wd) distinction.]

As for the fronting effects, inconsistency was found between the acoustic (F2) and the articulatory (x) dimensions [Figs. 2(a), 2(b)]. There was a significant fronting effect in F2 before a higher boundary (IP=ip>Wd, $p<0.01$), with no interaction with the Speaker. On the other hand, in the articulatory (x) dimension, there were no main effects but significant interactions with Speaker ($F[8,30]=12.23$, $p<0.001$ for TM; $F[8,30]=9.39$, $p<0.01$ for TD) such that only one speaker (S5) showed a significant fronting effect in TM.

For /ɑ♯/, there was significant lowering and backing in both the acoustic and the articulatory vowel spaces [Figs. 2(a), 2(c)]. In the acoustic space, boundary-induced lowering and backing of /ɑ♯/ was found in three speakers (S1, S2, S5 for lowering; S3, S4, S5 for backing). In the articulatory dimension, the lowering and backing effect becomes even clearer: four speakers (excluding S1) showed tongue lowering in either TM or TD and all five showed tongue backing before a higher prosodic boundary in both TD and TM.

In summary, in the vertical dimension, there was a consistent boundary-induced raising effect for /i♯/ in both acoustic and articulatory spaces; in the horizontal dimension, a fronting effect for /i♯/ was found primarily in the acoustic dimension. (See Sec. IV B for a discussion on this.) These

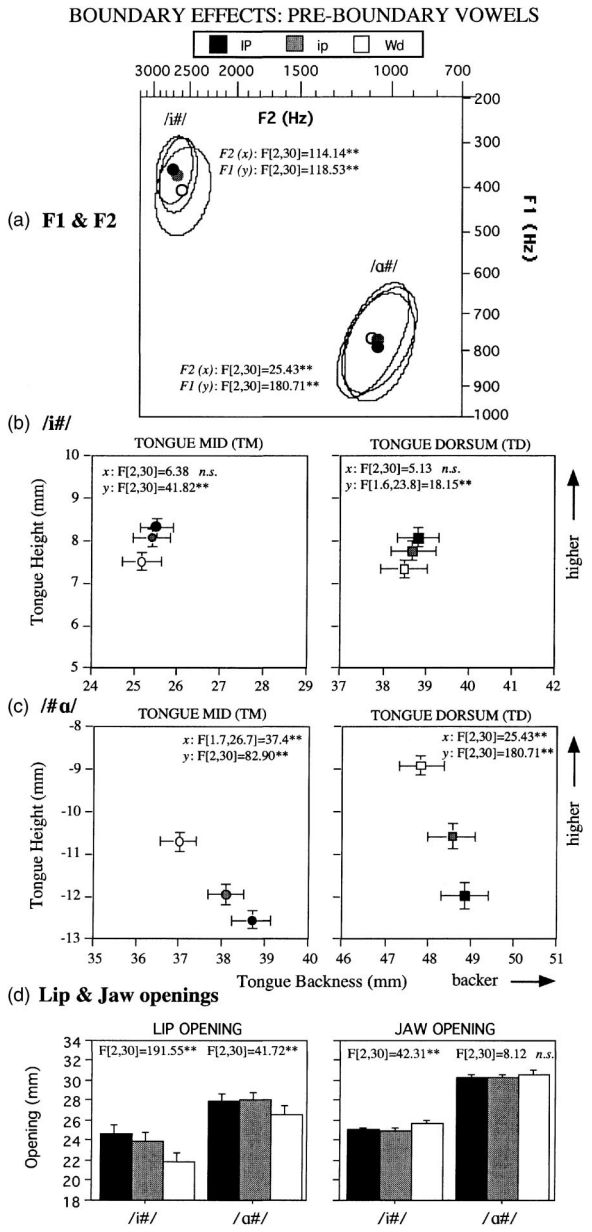


FIG. 2. Effect of Prosodic Boundary on F1 and F2 (a), the tongue maxima /i♯/ and /ɑ♯/ (b), (c), and the lip and jaw opening maxima (d) in the domain-final position. (* refers to $p<0.01$;** refers to $p<0.001$; $N=20$.) Note that in this case, N is 20 (5 speakers×2 accent conditions×2 accent conditions of the adjacent vowel.)

boundary effects on /i♯/ differed from the accent effects in that the former induced consistent acoustic/articulatory raising whereas the latter induced consistent acoustic/articulatory fronting. For the low/back vowel /ɑ♯/, there were consistent boundary-induced lowering and backing in both the acoustic and the articulatory vowel spaces. In particular, the boundary effect on backing of /ɑ♯/ was more consistent than was the accent effect, while both the boundary and the accent effects showed similar acoustic/articulatory lowering of /ɑ♯/.

b. Boundary effect on lip and jaw opening maxima in V_1

As shown in Fig. 2(d), the lip opening was significantly larger for both vowels before a higher prosodic boundary (IP=ip>Wd, $p<0.01$). There were no interactions with the Speaker. In contrast to the lip opening pattern, there was a symmetry in the jaw opening maxima between /i♯/ and /ɑ♯/. On the one hand, as shown in the right panel of Fig. 2(d), /i♯/ showed a small but significantly *smaller* jaw opening before a higher boundary (ip or IP) than before a lower one (Wd) ($p<0.01$), with no significant interaction with the Speaker. (All five speakers showed the same pattern.) On the other hand, /ɑ♯/ showed no main effect on the jaw opening but a significant interaction with the Speaker ($F[8,30]=11.98$, $p<0.01$). Notably, only one speaker (S5) showed an increase in the jaw opening at a higher boundary and two speakers (S2,S4) showed a decrease in the jaw opening at a higher boundary, as was the case for /i♯/.

In summary, domain-final vowels were produced with a larger lip opening before higher prosodic boundaries, but not necessarily with a larger jaw opening. In fact, for /i♯/, the jaw opening was reliably *smaller* at a higher boundary, as opposed to the accent-induced increase in the jaw opening for /i♯/. (See Sec. IV B for a relevant discussion on the discrepancy between the jaw and the lip openings.)

B. Articulatory maxima in domain-initial (post-boundary) vowels (♯CV_2)

1. Effect of accent on C_2V_2

a. Accent effect on F1 and F2 and tongue maxima in C_2V_2 For /♯bi/, there were consistent fronting effects, with no interactions with the Speaker, in both the acoustic and the articulatory dimensions [Figs. 3(a), 3(b)]. However, as was the case for the domain-final accent effect, there were unexpected accent-induced lowering effects in both the acoustic and articulatory dimensions; for the latter the effect was evident in TM. Only one speaker (S4) showed accent-induced raising for /i/ in both the acoustic and the articulatory vowel space. Conversely, four speakers (S1,S2,S3,S5) showed lowering in the acoustic dimension (F1), with two speakers (S1,S5) also showing lowering in the articulatory dimension.

For /♯bɑ/, accent entailed significant lowering and backing in the acoustic dimension with no interaction with the Speaker [Fig. 3(a)]. The articulatory data showed a corresponding tongue lowering effect [Fig. 3(c)], but not a corresponding backing effect. Only one speaker (S3) showed significant tongue backing and two (S2,S4) tongue fronting. (See Sec. IV A for possible explanations for the discrepancy between the acoustic and the articulatory data.)

In summary, as was for domain-final vowels, in the backness dimension domain-initial /i/ was consistently fronter when accented in both acoustic and articulatory spaces, but in the vertical dimension, it was actually *lower*, at least in the acoustic dimension. The low/back /ɑ/ was generally lower and backer when accented, although the backing effect for /ɑ/ was primarily evident in the acoustic dimension.

b. Accent effect on V*2 lip and jaw opening maxima* As shown in Fig. 3(d), accented V_2 is associated with greater jaw and lip openings than unaccented V_2, regardless of the vowel type. All speakers showed significantly greater jaw and lip opening patterns for accented domain-initial vowels.

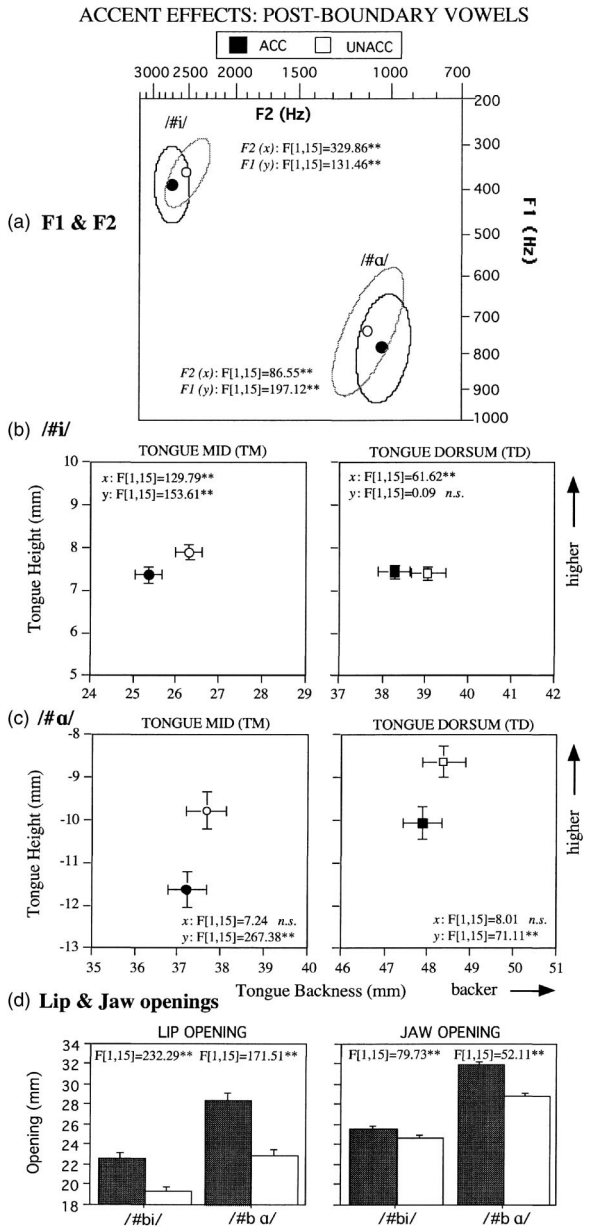


FIG. 3. Effect of Accent on F1 and F2 (a), the tongue maxima /♯i/ and /♯ɑ/ (b), (c), and the lip and jaw opening maxima (d) in the domain-initial position. (* refers to $p<0.01$;** refers to $p<0.001$; $N=30$.)

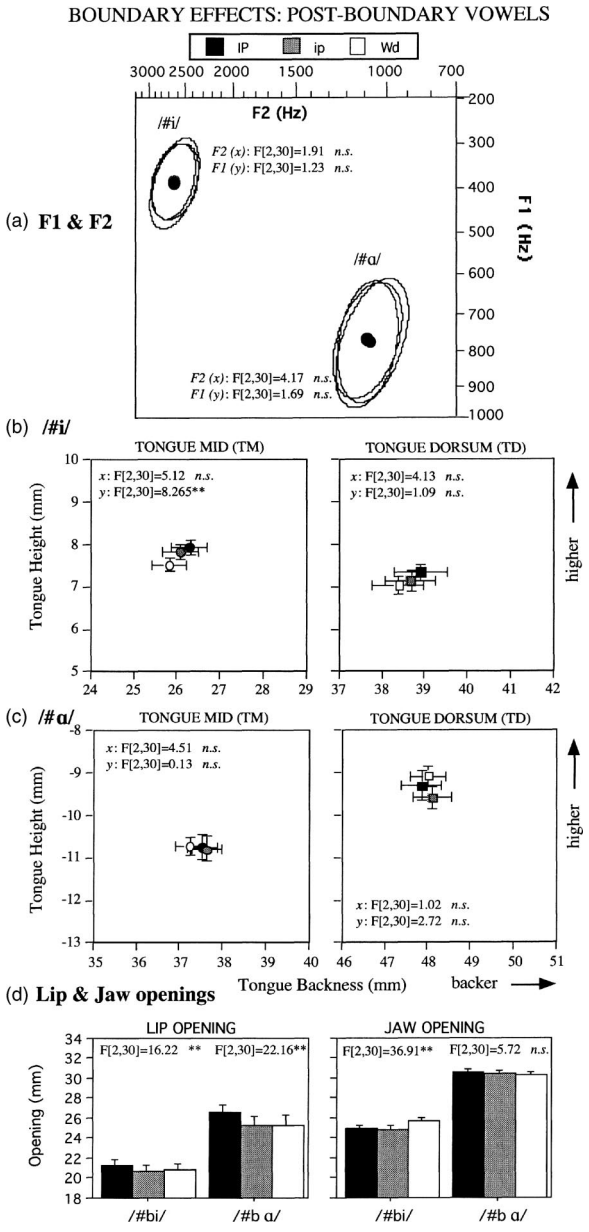


FIG. 4. Effect of Prosodic Boundary on F1 and F2 (a), the tongue maxima /♯i/ and /♯ɑ/ (b), (c), and the lip and jaw opening maxima (d) in the domain-initial position. (* refers to $p<0.01$;** refers to $p<0.001$; $N=20$.)

2. Effect of boundary type on domain-initial C_2V_2

a. Boundary effect on C_2V_2 *F1 and F2 and tongue maxima* The results with respect to the boundary effect on the domain-initial vowels are not as robust and consistent as the boundary effect on the domain-final vowels. Consider /♯bi/ in the acoustic F1–F2 vowel space [Fig. 4(a)]. There was no main effect of Boundary in the acoustic dimension, but significant interactions with Speaker ($F[8,30]=5.77$, $p<0.01$ in F1; $F[8,30]=6.68$, $p<0.01$ in F2). Only one speaker (S5) showed boundary-induced raising and backing (not fronting) for /i/, and all others showed no significant differences. On the other hand, in the articulatory dimensions, there was significant tongue raising for /i♯/ at a higher boundary (IP>Wd, $p<0.01$), as reflected in TM in Fig. 4(b). [Four speakers (S1,S2,S4,S5) showed a articulatory raising effect at a higher boundary, and one speaker (S3) showed the opposite trend.] By contrast, there was no evidence for boundary-induced tongue fronting: only one speaker (S5) showed a tongue fronting effect but two others (S3,S4) showed the opposite.

Turning to /♯bɑ/, as shown in Figs. 4(a), 4(c), both acoustic and articulatory data showed no main effect but significant interactions with the Speaker in all cases. In the backness dimension, two speakers (S3,S5) showed acoustic and articulatory backing for /ɑ/ after a higher prosodic boundary, and an additional speaker (S4) showed articulatory backing only. In the height dimension, only one speaker (S1) showed the lowering of /ɑ/ in the acoustic dimension and two (S3,S5) in the articulatory dimension. Interestingly, the two speakers (S3,S5) who showed tongue lowering were the ones who showed acoustic (and articulatory) backing for /ɑ/.

In summary, domain-initial /i/ was higher at a higher boundary, more so in the articulatory than in the acoustic dimension. One speaker (S5) showed both acoustic and ar-

ticulatory raising and three speakers (S1,S2,S4) showed articulatory raising only. Unlike the domain-final /i/, however, there was no consistent evidence of either an acoustic or an articulatory fronting effect: only S5 showed a fronting effect only in the articulatory dimension. The effect of Boundary on domain-initial /ɑ/ was less clear. Notably, however, two speakers (S3,S5) showed boundary-induced backing in both the acoustic and the articulatory dimensions, and the same two speakers also showed articulatory tongue lowering, which was not coupled with acoustic lowering. (See Sec. IV A for a discussion on this.)

b. Boundary effect on C_2V_2 *lip and jaw maxima* There was a significant main effect of Boundary on the maximum lip opening for /♯bi/ (IP>ip=Wd, $p<0.01$) seen in Fig. 4(d). But there was a significant interaction with Speaker ($F[8,30]=19.77$, $p<0.001$): two speakers (S1,S5) showed a significantly larger lip opening after a higher boundary (IP≥ip>Wd), while two others (S2,S4) showed nonsignificant trends in the same direction. The jaw opening data for /♯bi/ also showed a significant main effect, as shown in the right panel of Fig. 4(d), but in the opposite direction: the jaw opening was smaller for IP and ip than for Wd ($p<0.01$). Four speakers indeed showed a decreased jaw opening at a higher boundary (IP=ip<Wd, $p<0.01$). (Recall that the same four speakers showed a boundary-induced decrease in the jaw opening for the domain-final /i♯/.)

As for /♯bɑ/, there was a main effect of Boundary, showing a significantly larger lip opening at a higher prosodic boundary (IP>ip=Wd, $p<0.01$), as shown in Fig. 4(d). There was, however, a significant interaction with Speaker ($F[8,30]=11.21$, $p<0.01$) due to one speaker (S3), who showed the opposite pattern. (Four speakers showed a similar pattern, but only significantly so for S1, S5). For the jaw opening for /♯bɑ/, there were no effects of Boundary.

In summary, for domain-initial /i/, the lip opening is generally larger at a higher prosodic boundary (found in four speakers) but the jaw opening is *smaller* (found in four speakers) at higher prosodic boundaries; for domain-initial /ɑ/, the lip opening was generally larger at higher boundaries but there was no boundary effect at all on the jaw opening.

IV. SUMMARY AND DISCUSSION

The results are recapitulated below and in Table II.

A summary of accent effects (when accented) are as follows: (1) Both the lip and the jaw openings for both /ɑ/ vs /i/ are generally larger; (2) the high front vowel /i/ is consistently fronter but not necessarily higher (only two speakers showed raising effects only in the acoustic dimension); and (3) the low back vowel /ɑ/ is generally lower, but backer only in the acoustic vowel space.[5]

A summary of boundary effects (at a higher boundary) are as follows: (1) The lip opening for both vowels /i,ɑ/ is generally larger, but more consistently so in domain-final positions than in domain-initial ones; (2) the jaw opening is not larger. Instead, it is significantly and consistently smaller for /i/, while there is no consistent jaw opening pattern for /ɑ/; (3) the high front vowel /i/ is consistently higher in domain-final positions. But in domain-initial positions, it is generally higher only in the articulatory vowel space; (4) the high front vowel /i/ is consistently fronter in domain-final positions primarily in the acoustic space. In domain-initial positions, it is hardly fronter; (5) the low back vowel /ɑ/ is consistently lower in domain-final positions. In domain-initial positions, there is no significant lowering effect for /ɑ/; and (6) the low back vowel /ɑ/ is consistently backer, though the effect is more consistently so in domain-final than in domain-initial position.

TABLE II. A summary of the main effects of Accent and Boundary Type on domain-final (_♯) and domain-initial (♯_) vowels. Cells with descriptions indicate that the main effects are significant at $p<0.01$ with "*" and $p<0.001$ with "**." When there are significant interactions between a prosodic factor and the factor Speaker, the individual speaker's contribution to the observed pattern is also noted parenthetically based on the results of a series of *t*-tests conducted between conditions with each speaker.

V-type	Phonetic dimensions	Accent effect		Boundary effect	
		_♯ (V1)	♯_ (V2)	_♯ (V1)	♯_ (V2)
/i/	F1	*n.s.* H (Ss4,5) L (S1)	*n.s.* H (S2) L (Ss1,3,4,5)	** H	*n.s.* H (S5)
	Tongue *y*	** L (Ss1,2)	** H (S2) L (Ss1,5)	** H	** H (Ss1,2,4,5) L (S3)
	F2	** F	** F	** F	*n.s.* B (S5)
	Tongue *x*	** F	** F	*n.s.* F (S5) B (Ss1,4)	*n.s.* F (S5) B (Ss3,4)
	Lip	** LG	** LG	** LG	** LG (Ss1,5)
	Jaw	** LG	** LG	** SM	** SM (Ss1,2,3,5)
/ɑ/	F1	** L	** L	** L (Ss1,2,5)	*n.s.* L (S1) H (S2)
	Tongue *y*	** L	** L	** L	*n.s.* L (Ss3,5) H (Ss2,4)
	F2	** B (Ss1,3)	** B	** B (Ss3,4,5)	*n.s.* B (Ss3,5) F (S1)
	Tongue *x*	*n.s.*	*n.s.* B (S3) F (Ss2,4)	** B	** B (Ss3,4,5)
	Lip	** LG	** LG	** LG	** LG (Ss1,5) SM (S3)
	Jaw	** LG	** LG	** LG (S5) SM (Ss2,4)	*n.s.*

Note: H=higher, L=lower, F=fronter, B=backer, SM=smaller, LG=larger; Ss=Speakers. Numbers in parentheses indicate speaker IDs.

A. Accent and featural enhancement

It was hypothesized that under accent, the maximum position of the tongue should increase in a direction to enhance place features, whereas the jaw and lip openings should be expanded in order to increase sonority. As far as the opening of the vocal tract is concerned, the results showed that accented vowels are produced with an increase in jaw and lip openings for both vowels /ɑ,i/, which is taken to be an enhancement of the sonority feature.

As for place features [high, low, back], defined in the acoustic F1–F2 vowel space, the results showed that /ɑ/ is generally lower and backer when accented versus unac-

cented, which can be interpreted as an enhancement of both [+low] and [+back] for /ɑ/. For the high front vowel /i/, however, it is consistently fronter, but not necessarily higher, suggesting that accentuation induces an enhancement of [−back], but not a consistent enhancement of [+high] insofar as the features are defined in the acoustic vowel space. This suggests that not all nonsonority (place) features defined in the acoustic dimension are enhanced uniformly under accent.

Now, let us consider how the acoustic results can be linked to the articulatory results. For /ɑ/, the acoustic lowering effect (enhancement of [+low]) was further evident in the articulatory space. The enhancement of [+low] as reflected in the raised F1 may be attributable to the articulatory maneuvers involving not only the jaw and lip opening but also the tongue lowering, converging on sonority expansion (e.g., Harrington *et al.*, 2000; Erickson, 2002). On the other hand, the enhancement of [+back] evident in backing in the acoustic space (lowered F2) was not consistently accompanied by the tongue backing in the articulatory space, (while the tongue lowering was observed in all speakers). In the present study, the articulatory vowel space reflects a Cartesian vowel space defined by the occlusal plane. The acoustic F2 lowering may then not be directly translatable in terms of the tongue backing in an articulatory space defined this way. Instead, as Wood (1979, 1982) proposed, the observed F2 lowering for /ɑ/ may be better accounted for by an increase in the constriction degree in the low pharyngeal area in the vicinity of the epiglottis. It is then plausible that the tongue lowering in a Cartesian vowel space may indeed increase the constriction degree in the low pharyngeal area, which would result in backing in the acoustic space (lowered F2). According to the nomogram, based on a two-tube model (e.g., Fant 1960), it is predicted that F2 as a back-cavity resonance for /ɑ/ increases as the length of the back-cavity gets shorter (Stevens, 1989; Ladefoged, 1996; Johnson, 1997). This possibility is indeed further supported by the boundary effects on /ɑ/ in domain-initial positions. Recall that the two speakers (S3,S5) who showed accent-induced (articulatory) tongue lowering were the ones who also showed backing in the acoustic space (lowered F2) for /ɑ/, which suggests a close relationship between the articulatory tongue lowering and the backing in the acoustic space, at least in these two speakers.

Turning to /i/, there was a closer mapping between the acoustic and articulatory vowel spaces. The accent-induced fronting effect (enhancement of [−back]) in the acoustic space was further manifested in the tongue fronting in the articulatory *x* dimension, whereas the nonraising effect (thus no evidence of [+high] enhancement) in F1 was commensurate to the nonraising effect in the articulatory *y* dimension. In fact, two speakers showed accent-induced lowering effects in both the acoustic (F1) and the articulatory (TM, tongue mid) vowel spaces (and two others in the acoustic space only). The tongue lowering for /i/ in TM for these speakers might be interpreted as the entire tongue body being shifted forward along the arc of the palate (as evidenced in the tongue fronting), which may rotate the tongue midposition slightly downward. Alternatively, the lowering of TM for accented /i/ could be interpreted, not in terms of place feature enhancement, but simply as a byproduct due to the tongue shifting to achieve a proper constriction degree and location in the area of the palate. However, these two alternatives do not fully explain why there is also a corresponding acoustic lowering effect (F1 raising) in the acoustic dimension. Instead, the acoustic and articulatory lowering effects observed in some speakers may be interpreted as a result of the articulatory maneuver coupled with the jaw lowering (and the lip opening) to increase sonority.

These hypotheses lead to an interesting observation. In the height dimension in both acoustic and articulatory, when a vowel has a place feature whose enhancement is in direct conflict with sonority expansion (i.e., [+high] for /i/), the enhancement of that feature appears to be suppressed, as found in most speakers' data. What appears to be enhanced then is the [−back] feature evident in both the acoustic and articulatory fronting, which at least has no antagonistic effect on sonority expansion. On the other hand, when a vowel has place features whose enhancement is not in conflict with sonority expansion (in the case of [+low] and [+back] for /ɑ/), no such constraint applies. This observation is also supported by a finding reported by de Jong (1995a), who showed that the tongue position for English vowel /ʊ/ (with two place features [+high] and [−back]) is backer but not necessarily higher under accent.[3]

This account suggests that accent-driven hyperarticulation of vowels in English is achieved by an enhancement of some, but not all, place features, with consistent sonority expansion. If the ultimate goal of accent is to maximize phonemic contrast and ultimately lexical distinctions, in the spirit of the hyperarticulation hypothesis, such distinctions are not necessarily maximized by enhancement of all the distinctive place features, but may actually be best enhanced by making segments louder as a result of sonority expansion, and other featural enhancement may enhance the peripherality of the vowels not in all directions but in ways that do not conflict with sonority expansion. Similarly, Fowler (1995) suggests that that the phonetic goal of accent is to simply maximize prominence in order for listeners to understand better (*the global effect hypothesis*).

B. Prosodic boundary and featural enhancement

Another hypothesis set forth for this study was that the phonetic realization of vowels at domain edges is driven primarily by sonority expansion, and not by hyperarticulation, such that all three articulators (the tongue, the jaw, and the lips) converge on a larger vocal tract opening at higer prosodic boundaries, resulting in acoustic lowering (i.e., F1 raising). In this vein, it was also predicted that nonsonority (place) features would not be enhanced if the results would be antagonistic to the sonority expansion. However, the results revealed that although there is some evidence for boundary-induced sonority expansion, this is not the only effect seen at the boundaries.

The lip opening was found to be generally larger for both /i/ and /ɑ/ at a higher boundary in both the domain initial and final positions, suggesting that sonority expansion is associated with edges of prosodic domains. Interestingly, however, unlike the accent effect, neither /i/ nor /ɑ/ showed

any significant increase in the jaw opening at a higher boundary in both the domain initial and final positions. Instead, /i/ showed a significant *decrease* in jaw opening at a higher boundary.

With respect to the enhancement of place features, /ɑ/ showed lowering and backing effects in both the acoustic and the articulatory vowel spaces (more consistently so in domain-final than domain-initial position), which could be interpreted as the enhancement of [+low] and [+back] as well as sonority expansion. The vowel /i/ also showed acoustic raising and fronting and articulatory raising (but not consistent fronting). This again can be interpreted as the enhancement of [+high] and [−back] at a higher boundary. Thus, from these results we can postulate that it is not only sonority-related features but also nonsonority features that are enhanced at a strong prosodic boundary, even if the enhancement of such nonsonority features (e.g., [+high] for /i/) works antagonistically with sonority expansion. This suggests that sonority is not the only driving force for strengthening effects at prosodic boundaries, rejecting the hypothesis of boundary phenomena as being governed by sonority only (e.g., Farnetani and Vayra, 1996). [The tongue raising for English /i/ is compatible with French EPG data reported in Fougeron (2001), in that the French /i/ in the /p♯_p/ context tends to have a greater degree of linguopalatal constriction at a higher boundary.]

It is also noteworthy that the *decreased* jaw opening at a higher boundary for /i/ indicates that, if anything, there is an additional adverse effect against boundary-induced sonority expansion for /i/. This is especially interesting because there is a discrepancy between the lip and the jaw opening maxima: /i/ is generally associated with a larger lip opening but a smaller jaw opening. One possibility might have to do with competing articulatory goals in the production of domain-edge /i/: boundary-driven sonority expansion versus the enhancement of the proper constriction degree required for /i/. It is plausible that the lip opening satisfies boundary-driven sonority expansion while the jaw contributes to an enhancement of [+high] for /i/. (Recall that for the accent effect on /i/, there was neither acoustic nor tongue/jaw raising.) The reduced jaw opening may also have an effect of reinforcing the unroundedness of /i/ that may be responsible at least in part for the increased F2 for preboundary /i/.

Asymmetry in domain-initial versus domain-final prosodic strengthening. As summarized above, the boundary effect is generally more robust domain-finally than domain-initially. In almost all cases, the observed boundary phenomena were evident in more speakers domain-finally than domain-initially. The weaker domain-initial results are roughly consistent with the EPG data in Fougeron and Keating (1997), which showed that the vocal tract opening for /no/ (in reiterant speech), was larger for phrase-initial positions than for phrase-medial positions only for one speaker out of three. This is presumably because the effect of domain-initial strengthening is robust primarily on the domain-initial consonant in ♯CV, but the degree of strengthening may wane when it comes to the vowel articulation. (See Byrd, 2000; Byrd and Saltzman, 2003 for a possible explanation in the framework of a mass-spring gestural model.) This does not mean to say that there is no domain-initial strengthening effect on the vowel in ♯CV, but to say that the effect is less robust as compared to the domain-final articulation. [Note that Cho (in press) also showed boundary-induced strengthening in ♯CV lip opening kinematics.]

Returning to domain-final phenomena, it has been proposed that domain-final articulation is explained primarily by temporal expansion (e.g., Edwards *et al.*, 1991; Beckman *et al.*, 1992). However, as described above, the results of the present study do indicate that domain-final articulation is marked not only by temporal expansion but also by the enhancement of sonority and place features at a higher prosodic boundary—that is, by strengthening as well as lengthening. A similar domain-final strengthening effect on French /a/ was reported in Tabain (2003). However, the results of the present study are not fully comparable with previous studies. For example, consider the remark of Beckman *et al.* (1992) regarding domain-final articulation that ''[u]nlike the greater length of a nuclear-accented syllable (with a greater articulatory displacement), intonation-phrase-final lengthening is *not* accompanied by any significant difference in articulator displacement.'' The basis for this remark was that the jaw *closing* gesture for ''pop'' is temporally longer, but not spatially larger domain-finally. However, this remark does not seem to pertain to our tongue and lip opening data. One possible explanation is that the discrepancy comes from the difference in syllable structure as well as the articulatory parameters. The Beckman *et al*. conclusion was based on the V-to-C closing displacement data with closed syllables (/pɑp/) whereas the data in the present study were obtained from the articulatory (static) maxima with open syllables (/bɑ♯/). Another possibility is that the jaw may be less sensitive to changes in the level of prosodic boundary than other articulators. The data presented in this paper show that among the three articulators that were examined, it is only the jaw maxima that do not show expanded articulation for domain-final /ɑ/ (only one out of five speakers showed the jaw lowering effect), showing less sensitivity of the jaw to prosodic boundaries. This is consistent with Tabain's (2003) claim that the jaw is less sensitive to prosodic boundaries than is the tongue. Given these two possible explanations, the findings of the current study need not be seen as contradicting those of Beckman *et al.*, but as complementing them.

C. Boundary versus accent-induced strengthening

We have seen that vowels are more extremely articulated when they are either accented or located at a higher prosodic boundary, suggesting that there is some sort of prosodic strengthening. One of the major goals of this paper was to examine whether the prosodic strengthening that occurs at the edges of prosodic domains can be accounted for by the same principle that governs accent-induced strengthening.

The results suggest that the two aspects of prosodic structure are indeed differentiated by distinct phonetic patterns. One of the most comprehensible differences between the two effects comes from the way in which the high front vowel /i/ is produced. Accented /i/ was associated primarily with fronting in both the acoustic and the articulatory vowel spaces, which was accompanied by an increase in both the

jaw and the lip opening. On the other hand, domain-final /i/ at higher boundaries was associated primarily with raising in both the acoustic and the articulatory spaces, this time accompanied by a *decrease* in the jaw opening but an increase in the lip opening. We then infer that phonemic distinction is maximized primarily via enhancement of the frontedness ([−back]) under accent but via enhancement of the height ([+high]) at higher boundaries, while sonority is expanded in both contexts. For /ɑ/, the difference between the accent and the boundary effects was less clear, but the obvious difference was that while there was a clear jaw opening effect for accented /ɑ/, there was virtually no jaw lowering effect for domain-final /ɑ/. Despite all these differences, however, there is one common property that characterizes prosodic strengthening, regardless of whether it comes from accentuation or boundary: vowels are produced with an increase in sonority as evident in the lip opening, which is taken to be a consistent way of increasing linguistic prominence at prosodically strong locations.

D. Maintenance of phonological contrasts

Before concluding this paper, it is worth discussing how the results of the present study can be linked to current phonological theories with respect to prosodically strong positions. In the recent phonological literature, prosodically strong positions have been considered by some phonologists as ''privileged'' or ''licensing'' positions in which phonological contrasts are most often maintained, in which segments act as triggers of phonological modification such as vowel harmony, and in which they themselves often resist such a modification (e.g., Beckman, 1998; Steriade, 1999; Barnes, 2002). Although it is still controversial whether such a positional privilege is phonetically grounded (being attributable to the richness of the phonetic cues associated with that position) or structurally driven (being purely attributable to the position itself), what is clearer is that, cross-linguistically, phonological contrasts are likely to be maintained in prosodically strong positions. This phonological view of strong positions as licensing for phonological contrasts has been attested in many languages, and is indeed in line with the prosodic strengthening effects in English observed in the present study: the prosodically conditioned enhancement of vowel features can be seen as maximally maintaining phonological contrasts in prosodically strong positions. The propensity of contrast maintenance is also evident in V-to-V coarticulatory resistance in prosodically strong positions (Cho, 2004). Moreover, such phonological theories may be further enriched by taking into account two aspects of prosodic strengthening found in the present study: prosodic strengthening has different phonetic effects depending on prosodic position and the degree of strengthening at domain edges varies with the level of the prosodic boundary. With respect to the interaction between prosodic strengthening and language-specificity, Cho and McQueen (in press) showed that VOTs tend to be shortened in strong prosodic positions in Dutch, as opposed to those in similar strong positions in English, which suggests that prosodically driven enhancement of phonological contrast (e.g., voiced/voiced phonemic distinction) is further modulated by a language-specific feature system. These findings together suggest that the investigation of cross-linguistic phonological patterns is warranted, with special attention paid to whether positional effects on phonological alternations and contrast maintenance are further conditioned by multiply-layered higher-level prosodic structure.

E. Conclusion

The results of this study suggest that the complexity of prosodic structure reflected in accentuation and hierarchically nested prosodic constituents appear to be indeed differentiated by speakers, as evident in discrete prosodic strengthening phenomena in different prosodic contexts (in accented syllables versus at edges of prosodic domains; domain-initially versus domain-finally). Building on previous findings, this implies that not only suprasegmental features (duration, pitch, loudness) but also fine-grained segmental phonetic cues are all manipulated to signal the linguistic structure of prosody. Another implication of this study is that prosodic strengthening is not simply a low-level phonetic event but a complex linguistic phenomenon, giving rise to the enhancement of phonological features. This implies that prosodic strengthening is related with linguistic prominence such that it makes a sound more distinct from neighboring segments (syntagmatic enhancement) and/or makes the sound distinct from other contrastive sounds in the language (paradigmatic/phonemic enhancement). This, in turn, means that the phonological theory of positional strength or licensing may need to be further elaborated by taking a high-level prosodic structure into account.

ACKNOWLEDGMENTS

This work has been supported by the Max Planck Institute for Psycholinguistics, and the NSF doctoral improvement grant (BCS-0001716) to the author and Patricia Keating. The author would like to thank Pat Keating and Sun-Ah Jun for their comments and advice; Adam Albright, Peter Ladefoged, Bruce Hayes, Ken de Jong, and Elizabeth Johnson for valuable suggestions on an earlier version of this paper. This work has been greatly improved by comments and criticisms from Anders Löfqvist, Mary Beckman, and an anonymous reviewer. The articulatory data were originally collected for the author's 2001 UCLA Ph.D. dissertation, but have been completely reprocessed and reanalyzed along with additional acoustic data. Therefore, the results reported jere are different from those in the authors' dissertation.

[1] But, See Harrington *et al.* (2000) in which one speaker showed temporal sequencing in achieving the sonority expansion first and then the hyperaticulation of the vowel features of accented /i/ in Australian English.

[2] Out of 320 sentences that contained a phrase boundary IP or ip), 302 sentences (94.3%) reached agreement. Each speaker contributed 28 to 32 IP sentences and 29 to 31 ip sentences, showing a balanced distribution across boundary types. Furthermore, sentences were also distributed roughly equally across accent conditions: ACC#ACC /ɑ/, 36 sentences; /i/, 39 sentences), UNA#ACC /ɑ/, 38 sentences; /i/, 38 sentences), UNA#ACC /ɑ/, 36 sentences; /i/, 39 sentences) and UNA#UNA /ɑ/, 38 sentences; /i/, 38 sentences).

[3] 3This parallel should be taken with caution: The tongue height dimension in de Jong (1995a) was defined by a line of pellets on the lower molars

whereas the tongue height dimension in the present was defined by the occlusal plane.

Barnes, J. (**2002**). ''Positional neutralization: A phonologization approach to typological patterns,'' Ph.D. dissertation, University of California, Berkeley.

Beckman, J. N. (**1998**). ''Positional faithfulness,'' Ph.D. dissertation, University of Massachusetts, Amherst.

Beckman, M. E. (**1996**). ''The parsing of prosody,'' Lang. Cognit. Processes **11**, 17–67.

Beckman, M. E., and Ayers, G. (**1997**). ''Guidelines for ToBI labeling, version 3.0,'' unpublished Manuscript, Ohio State University.

Beckman, M. E., and Edwards, J. (**1994**). ''Articulatory evidence for differentiating stress categories,'' in *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, edited by P. A. Keating (Cambridge University Press, Cambridge), pp. 7–33.

Beckman, M. E., Edwards, J., and Fletcher, J. (**1992**). ''Prosodic structure and tempo in a sonority model of articulatory dynamics,'' in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. Docherty and D. R. Ladd (Cambridge University Press, Cambridge), pp. 68–86.

Beckman, M. E., and Pierrehumbert, J. (**1986**). ''Intonational structure in Japanese and English,'' Phonol. Yearbook **3**, 255–309.

Bolinger, D. (**1970**). ''Relative height,'' in *Prosodic Feature Analysis. Montreal*, edited by P. Léon, G. Faure, and A. Rigault (Marcel Didier), pp. 109–127 (reprinted in Bolinger (**1972**). *Intonation: Selected Readings* (Penguin Books, Harmondsworth).

Browman, C. P., and Goldstein, L. (**1992**). ''Articulatory phonology: an overview,'' Phonetica **49**, 155–180.

Byrd, D. (**2000**). ''Articulatory vowel lengthening and coordination at phrasal junctures,'' Phonetica **57**, 3–16.

Byrd, D., and Saltzman, E. (**2003**). ''The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening,'' J. Phonetics **31**, 149–180.

Cho, T. (**2002**). *The Effects of Prosody on Articulation in English* (Routledge, New York).

Cho, T. (**in press**). ''Manifestation of prosodic structure in articulation: Evidence from lip kinematics,'' in *Laboratory Phonology 8: Varieties of Phonological Competence*, edited by L. M. Goldstein, D. H. Whalen, and C. T. Best (Mouton de Gruyter, Berlin/New York).

Cho, T. (**2004**). ''Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English,'' J. Phonetics **32**, 141–176.

Cho, T., and Jun, J. (**2000**). ''Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean,'' in *Chicago Linguistics Society*, Vol. 36, No. 1, pp. 31–44.

Cho, T., and Keating, P. (**2001**). ''Articulatory and acoustic studies of domain-initial strengthening in Korean,'' J. Phonetics **29**, 155–190.

Cho, T., and McQueen, J. M. (**in press**). ''Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress,'' J. Phonetics **33**.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper & Row, New York).

de Jong, K. (**1995a**). ''The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation,'' J. Acoust. Soc. Am. **97**, 491–504.

de Jong, K. (**1995b**). ''On the status of redundant features: the case of backing and rounding in American English,'' in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge, UK), pp. 68–86.

Edwards, J. E., Beckman, M. E., and Fletcher, J. (**1991**). ''The articulatory kinematics of final lengthening,'' J. Acoust. Soc. Am. **89**, 369–382.

Edwards, J., and Beckman, M. E. (**1988**). ''Articulatory timing and the prosodic interpretation of syllable duration,'' Phonetica **45**, 156–174.

Erickson, D. (**2002**). ''Articulation of extreme formant patterns for emphasized vowels,'' Phonetica **59**, 134–149.

Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Mouton, The Hague).

Farnetani, E., and Vayra, M. (**1996**). ''The role of prosody in the shaping of articulation in Italian CV syllables,'' *Proceedings of the 1st ESCA Workshop on Speech Production Modeling*, 4th Speech Production Seminar, Autrans, France, pp. 9–12.

Fougeron, C. (**2001**). ''Articulatory properties of initial segments in several prosodic constituents in French,'' J. Phonetics **29**, 109–135.

Fougeron, C., and Keating, P. A. (**1997**). ''Articulatory strengthening at edges of prosodic domains,'' J. Acoust. Soc. Am. **101**, 3728–3740.

Fowler, C. A. (**1995**). ''Acoustic and kinematic correlates of contrastive stress accent in spoken English,'' in *Producing Speech: Contemporary Issues: For Katherine Safford Harris*, edited by F. Bell-Berti and J. J. Raphael, pp. 355–373.

Harrington, J., Fletcher, J., and Beckman, M. E. (**2000**). ''Manner and place conflicts in the articulation of accent in Australian English,'' in *Acquisition and the Lexicon: Papers in Laboratory Phonology V*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press, Cambridge), pp. 70–87.

Hoole, P. (**1996**). ''Issues in the acquisition, processing, reduction and parameterization of articulographic data,'' *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation*, München, Vol. 34, pp. 158–173.

Jackson, M. T. T. (**1988**). ''Analysis of tongue positions: Language-specific and cross-linguistic model,'' J. Acoust. Soc. Am. **84**, 124–143.

Johnson, K. (**1997**). *Acoustic and Auditory Phonetics* (Blackwell, Oxford).

Jun, S.-A. (**1993**). ''The phonetics and phonology of Korean prosody,'' Ph.D. dissertation, Ohio State University.

Jun, S.-A. (**1998**). ''The accentual phrase in the Korean prosodic hierarchy,'' Phonology **15**, 189–226.

Keating, P. A., Cho, T., Fougeron, C., and Hsu, C. (**1999**). ''Domain-initial strengthening in four languages,'' *UCLA Working Papers in Phonetics*, Vol. 97, pp. 139–151 (also to appear in *Papers in Laboratory Phonology 6, Cambridge* (Cambridge University Press, Cambridge).

Ladefoged, P. (**2000**). *A Course in Phonetics*, 4th ed. (Harcourt, Brace, Jovanovich, Fort Worth, TX).

Ladefoged, P. (**1996**). *Elements of Acoustic Phonetics*, 2nd ed. (The University of Chicago Press, Chicago).

Löfqvist, A. (**1999**). ''Interarticulator phasing, locus equations, and degree of coarticulation,'' J. Acoust. Soc. Am. **106**, 2022–2030.

Löfqvist, A., Gracco, V. L., and Nye, P. W. (**1993**). ''Recording speech movements using magnetometry: One laboratory's experience,'' *Forschungberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, Vol. 31, pp. 143–162.

Neary, T. N. (**1977**). *Phonetic Feature Systems for Vowels*, Ph.D. dissertation. University of Connecticut, Storrs, CT.

Pierrehumbert, J. (**1980**). ''The phonology and phonetics of English intonation,'' Ph.D. dissertation, MIT.

Pierrehumbert, J., and Beckman, M. E. (**1988**). *Japanese Tone Structure* (MIT Press, Cambridge, MA).

Selkirk, E. (**1986**). ''On derived domains in sentence phonology,'' Phonol. Yearbook **3**, 371–405.

Shattuck-Hufnagel, S., and Turk, A. E. (**1996**). ''A prosody tutorial for investigators of auditory sentence processing,'' J. Psycholinguist. Res. **25**, 193–247.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (**1992**). ''TOBI: A standard for labeling English prosody,'' in *Proceedings of the 1992 International Conference on Spoken Language Processing*, Vol. 2, pp. 867–870.

Steriade, D. (**1999**). ''Phonetics in phonology: The case of laryngeal neutralization,'' in *UCLA Working Papers in Linguistics*, No. 2, *Papers in Phonology 3*, edited by M. Gordon, pp. 25–146.

Stevens, K. N. (**1989**). ''On the quantal nature of speech,'' J. Phonetics **17**, 3–45.

Tabain, M. (**2003**). ''Effects of prosodic boundary on /aC/ sequences: articulatory results,'' J. Acoust. Soc. Am. **113**, 2834–2849.

Turk, A. E., and Shattuck-Hufnagel, S. (**2000**). ''Word-boundary-related durational patterns in English,'' J. Phonetics **28**, 397–440.

Westbury, J. R. (**1994**). ''On coordinate systems and the representation of articulatory movements,'' J. Acoust. Soc. Am. **95**, 2271–2273.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (**1992**). ''Segmental durations in the vicinity of prosodic phrase boundaries,'' J. Acoust. Soc. Am. **91**, 1707–1717.

Wood, S. (**1979**). ''A radiographic analysis of constriction locations for vowels,'' J. Phonetics **7**, 25–43.

Wood, S. (**1982**). ''X-ray and model studies of vowel articulation,'' Ph.D. dissertation, Lund University (see also Working Papers, Vol. 23, pp. 1–192, Dept. of Linguistics and Phonetics, Lund University).

The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy

Huei-Mei Liu[a)]
Department of Special Education, National Taiwan Normal University, Taiwan 98195

Feng-Ming Tsao
Department of Psychology, National Taiwan University, Taiwan

Patricia K. Kuhl
The Institute for Learning and Brain Sciences, University of Washington, Seattle, Washington

(Received 24 April 2004; revised 7 March 2005; accepted 8 March 2005)

The purpose of this study was to examine the effect of reduced vowel working space on dysarthric talkers' speech intelligibility using both acoustic and perceptual approaches. In experiment 1, the acoustic–perceptual relationship between vowel working space area and speech intelligibility was examined in Mandarin-speaking young adults with cerebral palsy. Subjects read aloud 18 bisyllabic words containing the vowels /i/, /a/, and /u/ using their normal speaking rate. Each talker's words were identified by three normal listeners. The percentage of correct vowel and word identification were calculated as vowel intelligibility and word intelligibility, respectively. Results revealed that talkers with cerebral palsy exhibited smaller vowel working space areas compared to ten age-matched controls. The vowel working space area was significantly correlated with vowel intelligibility ($r=0.632$, $p<0.005$) and with word intelligibility ($r=0.684$, $p<0.005$). Experiment 2 examined whether tokens of expanded vowel working spaces were perceived as better vowel exemplars and represented with greater perceptual spaces than tokens of reduced vowel working spaces. The results of the perceptual experiment support this prediction. The distorted vowels of talkers with cerebral palsy compose a smaller acoustic space that results in shrunken intervowel perceptual distances for listeners.
[DOI: 10.1121/1.1898623]

PACS numbers: 43.70.Dn, 43.70.−h, 43.71.Gv [RLD] Pages: 3879–3889

I. INTRODUCTION

A combination of both perceptual and acoustic analyses has revealed that using phonetic contrasts to explore dysarthric speech is a valid, sensitive, and reliable method to identify the critical acoustic-phonetic distortions that contribute to speech intelligibility deficits (Kent *et al.,* 1989; Kent *et al.,* 1997; Weismer and Martin, 1992; Yorkston, Dowden, and Beukelman, 1992). This acoustic-perceptual approach not only assesses overall communicative efficiency but also analyzes the characteristics of dysarthric speech to point out underlying speech-motor movement restrictions that are not easily identified from perceptual judgments. This study aimed to analyze the effect of restricted vowel acoustic space on speech intelligibility.

Cerebral palsy, a nonprogressive neuromuscular disorder, is usually associated with a variety of motor speech problems classified in a generic category, dysarthria (Hardy, 1983; McDonald, 1987). Approximately 30%–90% of talkers with cerebral palsy have some dysarthric speech and reduced speech intelligibility (Yorkston, Beukelman, and Bell, 1988). Disturbances in articulatory coordination of the lips, tongue, mandible, and velum in talkers with cerebral palsy have been reported (Kent, Netsell, and Abbs, 1978). These distorted physiologic mechanisms of speech appear to adversely affect intelligibility.

Perceptual studies using minimal phonetic contrasts to assess the relative contributions of individual phonetic cues to overall speech intelligibility have shown that vowel clarity can be a very potent predictor of speech intelligibility for dysarthric speech (e.g., Kent *et al.,* 1989; Weismer and Martin, 1992). For example, the clarity of high–low, front–back, and long–short vowel contrasts are highly correlated with word intelligibility (Ansel and Kent, 1992; Liu, Tseng, and Tsao, 2000; Whitehill and Ciocca, 2000). Though vowel errors can decrease speech intelligibility, they seem less severe when compared to more problematic consonant contrasts (Platt, Andrews, and Howie, 1980; Platt *et al.,* 1980).

Furthermore, the important acoustic cues for distinguishing minimal contrasts among vowels, such as duration, first formant ($F1$), second formant ($F2$), and $F2-F1$ differences, were also shown to account for much of the variance in word intelligibility (Bunton and Weismer, 2001; Turner, Tjaden, and Weismer, 1995; Weismer *et al.,* 2001). This suggested that distorted acoustic features for vowels, such as extremely long duration, distorted formant location, and atypical formant slope, were highly associated with the speech intelligibility deficits in speakers with dysarthria.

Other studies analyzing dysarthric speech have also found that distorted vowels adversely affect intelligibility

a)Electronic mail: liumei@cc.ntnu.edu.tw

(Ansel and Kent, 1992; Liu *et al.*, 2000; Whitehill and Ciocca, 2000). Ansel and Kent (1992) investigated the speech intelligibility of 16 male adults diagnosed with mixed-type cerebral palsy, showing that the acoustic correlates of all vowel contrasts tested (i.e., high–low, front–back, and lax–tense) significantly contributed to overall speech intelligibility. Liu and colleagues (2000) assessed the weightings of six minimal phonetic contrasts on speech intelligibility of Mandarin-speaking young adults with cerebral palsy. They found that the segmental intelligibility of a front–back vowel contrast and its acoustic correlate (i.e., the formant frequency difference between $F1$ and $F2$) contributed greatly to overall speech intelligibility. In addition, Whitehill and Ciocca (2000) used principal component analysis and found that vowel contrasts were combined as an important component accounting for the total variance of intelligibility. These acoustic and perceptual findings on speech intelligibility of talkers with cerebral palsy have consistently demonstrated that imprecise vowel articulation greatly reduced intelligibility.

A. Vowel working space and vowel intelligibility

Vowels are quasiperiodic sounds, and the formant pattern of a given vowel determines its phonetic quality and reflects the articulatory configuration that produced it (Stevens, 1998). Generally, $F1$ varies mostly with tongue height such that the higher the tongue position, the lower the $F1$ frequency, while $F2$ varies mostly with tongue advancement; the more anterior the tongue position, the higher the $F2$ and larger $F2$ to $F1$ difference (Kent, Dembowski, and Lass, 1995; Stevens, 1998).

Vowel working space area can be seen as an index of the accuracy of vowel articulation, which signifies gross motor control ability of the tongue and jaw coordination. To examine the vowel working spaces of individual talkers, the ''corner'' vowels, such as /i/, /a/, and /u/, are frequently selected because these vowels are the most common in human languages (Ladefoged and Maddieson, 1996). In addition, these vowels are perceptually and acoustically exceptional because they represent the extreme positions in a talker's articulatory vowel working space, and hence extreme formant frequency values in acoustic space (Lindblom, 1990).

Restricted size of the vowel working space enclosed by the first two formants of three or four corner vowels has been shown in dysarthric speech for talkers with amyotrophic lateral sclerosis (ALS) (Weismer *et al.*, 1992; 2001), traumatic brain injury (Ziegler and von Cramon, 1983, 1986), and even for children with dysarthria (Hodge, 1999; Higgins and Hodge, 2001). Moreover, studies have reported a positive relationship between speech intelligibility and the area of vowel working space formed by the corner vowels /i/, /æ/, /a/, and /u/ on the $F1/F2$ coordinates for English-speaking adults with dysarthric speech (Turner *et al.*, 1995; Weismer *et al.*, 2000).

In addition to atypical talkers, normal American talkers who produced larger vowel working spaces were generally judged to be more intelligible than talkers with smaller spaces (Bradlow, Torretta, and Pisoni, 1996). Relatively expanded vowel working space was observed for clear speech compared to conservational speech, while speaking rate remained similar (Krause and Braida, 2004). Moreover, recent acoustic-phonetic work on assessing vowel exaggeration in caregivers' speech directed to infants demonstrated that the vowel working space was significantly larger in infant-directed speech than in adult-directed speech (Kuhl *et al.*, 1997; Burnham, Kitamura, and Vollmer-Conna, 2002; Liu, Kuhl, and Tsao, 2003). Thus, the area of vowel working space may affect perceived speech intelligibility and could serve as a measure of talker variability. In other words, vowel working space area may indicate speech intelligibility and the overall severity of speech impairment for talkers with dysarthria.

Although many studies have assessed vowel working space area for American English-speaking adults with dysarthria, there is a paucity of literature on Mandarin talkers with dysarthria. Mandarin Chinese is a tonal language in which every syllable must carry a lexical tone, a suprasegmental element that serves to signify meaning. The primary acoustic correlate of tones is fundamental frequency ($F0$). Compared to 14 vowels of American English, there are 8 different vowels of Mandarin Chinese. Although the numbers of vowels vary between languages, the vowel system of Mandarin Chinese, similar to most of the vowel systems in the world, includes high and low, front and back, and rounded and unrounded contrasts (Howie, 1976; Tseng, 1990). To test the universality of restricted vowel working space on speech intelligibility, the systematic investigation of acoustic-perceptual association in dysarthric speech must be extended to more languages. The target language of this study, Mandarin Chinese, is currently spoken by more than one billion people. Despite the widespread use of Mandarin Chinese, there has been a very limited amount of work of segmental acoustics in adults with cerebral palsy in this language (Liu *et al.*, 2000).

The first experiment of this study examined the relationship between vowel working space area and overall speech intelligibility in Mandarin talkers with cerebral palsy, to examine how well the data from English talkers with dysarthria generalize to another language population.

B. Vowel working space, perceptual organization, and speech intelligibility

The acoustic-perceptual analysis of the relationship between vowel working space area and speech intelligibility sheds light on how an individual talker's reduced vowel working space can affect listeners' identification of speech sounds. However, correlational analyses may not be able to fully determine the effect of reduced vowel working space on speech intelligibility. Additional speech features, such as breathy voice, inconsistent vocal intensity, and unstable speaking rate of vowels (e.g., Kent *et al.*, 1978) common in dysarthric speech can combine with restricted vowel working space to reduce speech intelligibility. Besides, the perceptual organization of listeners perceiving distorted acoustic signals with reduced vowel working space is unclear. The perceptual organization of speech sounds is an essential element of phonetic representation and crucial to the processing of acoustic input (Eimas and Miller, 1992; Remez, 2001). Studies have

shown that the perception of phonetic category goodness affects internal phonetic organization when listeners were engaged in speech judgments. For example, perceptual distance shrinks around good exemplars but expands around poor exemplars when physical distances are kept constant among tokens (Iverson and Kuhl, 1996; Iverson *et al.*, 2003). It is reasonable to hypothesize that the distorted vowels in dysarthric speech may occupy less dispersed perceptual space than typical vowels. In other words, speech sounds from talkers with reduced vowel working space may be associated with a smaller perceptual space, resulting in a negative impact on the judgment of speech intelligibility.

Few acoustic-phonetic studies on dysarthric speech have attempted to assess the effect of reduced vowel working space on listeners' perceptual organization of speech sounds. Determining the specific features that contribute to decreased speech intelligibility has not only theoretical value for the field of speech perception, but also has clinical implications. The combination of both acoustic and perceptual analysis should provide a better understanding of the causes of reduced speech intelligibility and could facilitate clinical interventions to improve intelligibility in this population. To further evaluate the role of vowel working space area on speech intelligibility, this study also systematically manipulated the vowel working space area to assess its effect on the perceptual organization of speech sounds.

The goal of the present study was to examine the effect of reduced vowel working space on dysarthric talkers' speech intelligibility using both acoustic-phonetic and perceptual approaches. Two related experiments were conducted. In experiment 1, the relationship between the area of vowel acoustic space and speech intelligibility was examined in 20 Mandarin-speaking young adults with cerebral palsy. A positive relationship between the enclosed area of corner vowels in the $F1/F2$ space and speech intelligibility should support a generalization of the findings of previous studies on English-speaking talkers with dysarthria to other languages. In experiment 2, vowel working space area was systematically manipulated to further examine the impact of reduced vowel working space area on overall speech intelligibility. To support the hypothesis that reduced vowel working space negatively affects the speech intelligibility, poor vowel exemplars (i.e., tokens of a relatively reduced vowel working space) are expected to be represented with smaller perceptual spaces than those of good exemplars (i.e., tokens of a relatively larger vowel working space).

II. EXPERIMENT 1: THE RELATIONSHIP BETWEEN VOWEL WORKING SPACE AND SPEECH INTELLIGIBILITY

A. Methods

1. Participants

The talkers were 20 male young adults referred by school speech-language pathologists and diagnosed with cerebral palsy and various degrees of speech deficits. Their ages ranged from 17 years to 22 years old ($M = 18.5$ years). Each talker was judged as having adequate intelligence and hearing ability to perform the speech recording task based on health records and experimenter's observation. In addition, data were collected from ten age-matched neurologically intact controls who participated in the same speech recording task.

2. Speech materials

A total of 18 (3 vowels$\times$6 words) high-frequency bisyllabic words was chosen from a Chinese word database (Wu and Liu, 1987) for speech recording and acoustic analysis. The frequencies of these Mandarin Chinese bisyllabic words were controlled. They contained the target vowels, /i/, /a/, and /u/, in the first syllable. The bisyllabic target words were constructed as C**V̲**CV(V), in which the syllable-final vowel could be a monophthong or a diphthong and the tone numeral was attached to the end of each syllable, such as [t**i̲**4tau4] vs. [t**a̲**4tau4]. These 18 target words can be grouped into to three minimal-phonemic word pairs for each vowel contrast, /i/–/a/, /u/–/a/, and /i/–/u/, to better control for neighboring phonemic environment. The point vowel, /i/, /a/, or /u/, in the first syllable of each word was the target vowel for acoustic analysis. These 18 vowel-contrast words were taken from a Mandarin word intelligibility test used in a previous study (Liu *et al.*, 2000).

3. Recording procedures

During the recording, talkers were seated in a quiet recording room. Each of the 18 words was printed individually on a card. Before the recording, subjects were instructed to go through the 18 words to make sure that they knew the words. The presentation sequences of words were block randomized across talkers to decrease the talker's awareness of phonemic contrasts during speech recording. Talkers were asked to read each word aloud using a normal speaking rate in a comfortable situation. To control the speaking rate of individual talkers, an experimenter turned over each printed card for the talker, inducing an interword interval of about 3 s. Speech samples were tape recorded on a cassette recorder (Nakamichi 550). These speech samples were used as stimuli for both intelligibility judgments and acoustic analysis.

4. Speech intelligibility measures

An item identification task was employed for measuring speech intelligibility. Sixty listeners used Chinese characters to orthographically transcribe speech samples produced by 20 talkers with cerebral palsy. To control for the influence of familiarity with dysarthric speech on intelligibility judgments, none of the listeners was familiar with dysarthric speech. Three listeners simultaneously transcribed speech samples of one talker and none of these listeners heard more than one talker. In total, 60 (3×20) undergraduate students with normal hearing served as judges for the subjects with cerebral palsy. For the ten normal speakers, 30 (3×10) undergraduate students with normal hearing served as judges.

Two intelligibility scores were calculated based on the correct identification of the phonemes within 18 bisyllabic words for individual talkers. The percentage of correct vowel identifications was used as a measure of vowel intelligibility.

Likewise, the percent correct of the word identifications, i.e., correct identification of all phonemes of the bisyllabic word, was calculated as word intelligibility. The intelligibility scores, including the vowel intelligibility and word intelligibility, for each talker were averaged from the scores of the three judges. Given that two intelligibility scores were collected, the word intelligibility score was used to represent individual talkers' speech intelligibility in this study.

In addition to identifying the percentage of correct phonemes for one individual talker, all 60 listeners also transcribed a subset of the same three words randomly selected from one talker with cerebral palsy whose word intelligibility was about 50%. This was done to assess interjudge reliability, which was relatively high, Cronbach $\alpha = 0.95$. The same interjudge reliability procedure was utilized to assess the reliability for normal speakers. The reliability was also relatively high, Cronbach $\alpha = 0.997$.

5. Acoustic analysis procedures

The first and second formant frequencies ($F1$ and $F2$) of corner vowels (/i/, /a/, and /u/) contained in 18 target words produced by talkers with cerebral-palsy and normal talkers were measured. All formant measurements were conducted using a 486 DX2-66 PC with a speech analyzer, the Kay Elemetrics' Computerized Speech Laboratory (CSL) Software CSL-4300B. The target words were low-pass filtered at 8 kHz and sampled at a rate of 20 kHz. Both the linear prediction coding (LPC) spectra and wideband spectrograms were used to determine the locations of first two formants. The onset of each vowel segment was marked when both $F1$ and $F2$ became visible on the spectrogram. Vowel offset was marked at a point where $F2$ and/or $F1$ were no longer visible. A third marker was placed at the durational center of each vowel segment. Vowel formant frequencies were measured at the cursor that marked the central position of a vowel segment. If the LPC method did not clearly locate the vowel formants (because of vocal irregularities), the fast Fourier transform (FFT) spectra were derived as a supplementary analysis of vowel spectra.

After obtaining the formant frequencies of the three corner vowels /i/, /a/, and /u/, average $F1$ and $F2$ values for each vowel were plotted on an $F1/F2$ plane to provide quantitative indices of ''acoustic vowel working space area'' of individual talkers. The $F1$ and $F2$ pairs of each vowel were viewed as coordinates in the $x-y$ plane and the area was calculated using the following equation:

$$\text{Vowel triangle area} = \text{ABS}\{[F1\text{i}^*(F2\text{a}-F2\text{u}) + F1\text{a}^*(F2\text{u}-F2\text{i}) + F1\text{u}^*(F2\text{i}-F2\text{a})]/2\},$$

where ''ABS'' is absolute value, ''$F1$i'' symbolizes the $F1$ value of vowel /i/, ''$F2$a'' symbolizes the $F2$ value of vowel /a/,..., and so on.

Interjudge reliability of acoustic analysis was controlled in two ways. First, before analyzing the dysarthric speech samples, reliability between the experimenter and another person who was proficient in acoustic analysis was calculated on 10% of a normal talker's vowel samples. The interjudge reliability of this acoustic analysis procedure was 0.94. The average differences between judges were 25 Hz (4.6%) and 84 Hz (4.1%) for $F1$ and $F2$, respectively. Second, 10% of the measurements of the dysarthric vowel samples were used to calculate the interjudge reliability. The interjudge reliability of this acoustic analysis procedure was 0.91. The average differences between judges were 41 Hz (5.9%) and 103 Hz (5.1%) for $F1$ and $F2$, respectively.

B. Results and discussions

1. Speech intelligibility scores

Intelligibility scores of talkers with cerebral palsy are reported in Table I. The intelligibility scores of 20 talkers with cerebral palsy obtained from the item-identification task varied greatly across individuals (for word intelligibility, ranged from 24% to 88%, $M = 56\%$, SD $= 20.75$; for vowel intelligibility, ranged from 35% to 100%, $M = 78\%$, SD $= 17.87$). This result indicates that talkers in this study represent a wide-range severity of speech impairment. Compared to the great variation among individuals with cerebral palsy, normal speakers' intelligibility scores were in a narrower range (for word intelligibility, $M = 96.6\%$, SD $= 2.02$, range $= 92.2\% - 99.3\%$; for vowel intelligibility, $M = 99.8\%$, SD $= 0.59$, range $= 98.1\% - 100\%$). In addition, across all 20 talkers with cerebral palsy, the vowel intelligibility scores were highly correlated with word intelligibility ($r = 0.86$, $p < 0.0001$).

2. Vowel working space characteristics

Results of the vowel working space area encompassing /i/, /a/, and /u/ of individual talkers show that 20 talkers with cerebral palsy exhibited significantly smaller vowel working space areas when compared to ten age-matched controls, $F(1,28) = 13.025$, $p < 0.001$, Hedges's $g = 1.36$. Figure 1 displays more centralized vowel dispersions for talkers with cerebral palsy compared to those in the normal control group. The reduced vowel working space area represents a restriction of tongue elevation and front–back movement for talkers with cerebral palsy. Statistical analyses on vowel formants provide further support.

For the talkers with cerebral palsy, the low vowel /a/ exhibited significantly lower $F1$ frequency ($M = 729$ Hz, SD $= 109$ Hz) than was found for normal talkers [$M = 803$ Hz, SD $= 55$ Hz $F(1,28) = 3.98$, $p < 0.01$, Hedges's $g = 0.75$]. Conversely, the average $F1$ value of the high vowel /i/ in talkers with cerebral palsy ($M = 313$ H, SD $= 31$ Hz) was significantly higher than that of normal talkers [$M = 277$ Hz, SD $= 24$ Hz $F(1,28) = 8.85$, $p < 0.01$, Hedges's $g = 1.12$]. $F1$ variation is primarily associated with tongue height during vowel production (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). The narrower range of $F1$ values for the high-low vowel contrast (/i/–/a/), $F(1,28) = 8.11$, $p < 0.01$, Hedges's $g = 1.07$, implies that talkers with cerebral palsy have more limited jaw/tongue vertical movement compared to normal talkers.

In vowels from talkers with cerebral palsy, the $F2-F1$ difference was 1844 Hz (SD $= 190$) for the high-front vowel /i/ and it was 569 Hz (SD $= 93$) for the high-back vowel /u/.

TABLE I. Descriptive data for 20 male cerebral-palsied subjects.

Subject	Age	Classification	Severity of impairment	Word intelligibility (%)	Vowel intelligibility (%)	Vowel space area (Hz^2)
CP1	19	Athetoid	Moderate	59	69	205 539
CP2	19	Spastic	Moderate	24	35	164 606
CP3	19	Spastic	Moderate	36	57	238 559
CP4	17	Ataxic	Moderate	74	85	348 575
CP5	18	Tremor	Moderate	32	76	291 836
CP6	19	Mixed	Moderate	26	54	115 005
CP7	22	Mixed	Severe	35	63	111 157
CP 8	18	Spastic	Moderate	83	93	376 995
CP 9	20	Spastic	Moderate	84	98	353 263
CP 10	18	Spastic	Mild	55	81	125 456
CP 11	17	Spastic	Mild	88	100	296 440
CP 12	18	Athetoid	Moderate	53	91	269 073
CP 13	18	Athetoid	Mild	50	83	129 092
CP 14	19	Athetoid	Moderate	56	65	193 306
CP 15	18	Spastic	Moderate	44	78	280 635
CP 16	20	Spastic	Severe	74	96	362 846
CP 17	18	Athetoid	Severe	44	67	145 937
CP 18	18	Tremor	Moderate	84	96	396 053
CP 19	18	Athetoid	Severe	45	69	287 416
CP 20	17	Spastic	Moderate	75	100	228 411

When compared to normal talkers (/i/, M=2075 Hz, SD=132; and /u/, M=478 Hz, SD=39), talkers with cerebral palsy demonstrated a significantly narrower range of $F2-F1$ values for the front–back vowel contrast (/i/–/u/), $F(1,28)=15.18$, $p<0.001$, Hedges's $g=1.47$. Given that $F2$ or $F2-F1$ difference varies mostly with tongue advancement, the more anterior the tongue position, the higher the $F2$ and larger $F2-F1$ difference; the reduced range of $F2-F1$ indicates more restricted front–back tongue movement (Kent, Dembowski, and Lass, 1995; Stevens, 1998). Taken together, the smaller vowel working space and more centralized articulation reflect more restricted vertical and horizontal tongue movements in talkers with cerebral palsy compared to normal talkers during vowel production.

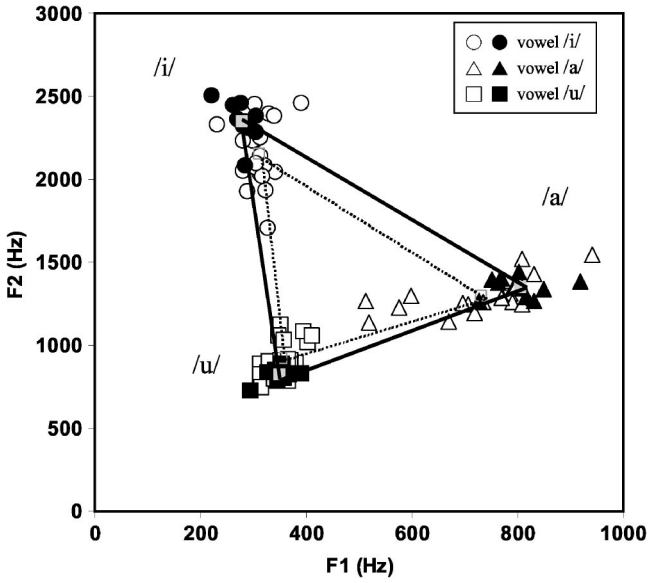


FIG. 1. Vowel working spaces of Mandarin talkers with normal speech and cerebral palsy. Vowel working spaces formed by the average of the $F1$ and $F2$ values for the vowels /i/, /a/, and /u/. Data points represent the mean coordinate of the first two formant frequencies of the vowels /i/, /a/, and /u/ produced by normal talkers (closed) versus talkers with cerebral palsy (open). Each data point represents the average of one talker's vowel productions. A decrease of the vowel working space is observed in talkers with cerebral palsy (dashed line).

3. The relationship between vowel working space area and speech intelligibility

To generalize the previous findings on the relationship between articulation space and speech intelligibility, the correlational analysis for Mandarin talkers with cerebral palsy was expected to demonstrate positive correlations between vowel working space area and speech intelligibility. The results of the analysis support this prediction. Vowel working space area was significantly correlated with the vowel intelligibility ($r=0.632$, $p<0.005$) and with word intelligibility ($r=0.684$, $p<0.005$).

In experiment 1, talkers with relatively larger vowel working spaces were judged to be more intelligible than talkers with reduced vowel working spaces. Therefore, vowel working space area could be a general indicator of speech intelligibility. Larger vowel working spaces have been shown to correlate with better speech intelligibility in typical talkers (Bradlow *et al.*, 1996). Also, smaller vowel working space area is commonly found in English speakers with dysarthria, secondary to disorders such as amyotrophic lateral sclerosis (ALS) (Turner *et al.*, 1995; Weismer *et al.*, 2001) and traumatic brain injury (Ziegler and von Cramon, 1983, 1986). These results of acoustic analysis on Mandarin talkers with cerebral palsy suggest that vowel working space area is important in the prediction of intelligibility in talkers with dysarthria of this type.

III. EXPERIMENT 2: EFFECTS OF REDUCED VOWEL WORKING SPACE ON LISTENERS' PERCEPTUAL ORGANIZATION

Speech intelligibility is a global measure, reflecting the cumulative effects of various acoustic-phonetic features on listeners' perceptual judgments. In experiment 1, results of the acoustic analysis showed that vowel working space area is strongly associated with speech intelligibility for talkers with cerebral palsy. Experiment 2 was designed to examine the effects of reduced acoustic vowel working space on normal listeners' perceptual organization through the manipulation of vowel working space area, while controlling for other speech characteristics across vowel tokens.

The stimulus set for this experiment consisted of synthesized /i/, /a/, and /u/ vowels varying in $F1$ and $F2$ frequencies. Six vowel triangles were generated with similar shape but different areas. The largest of these variants created vowel working spaces that corresponded to actual vowel working spaces of the most intelligible dysarthric speech in experiment 1. The smallest vowel working space was even smaller than that of the least intelligible talker in the previous experiment. During the experimental session, participants identified each token as /i/, /a/, or /u/, and rated the category goodness of each token on a 7-point scale. Subjects then completed an AX discrimination task in which they judged whether the tokens in each pair were the same or different.

Reaction times in the discrimination task were analyzed by multidimensional scaling (MDS) to map the perceptual organization underlying these stimuli and illustrate the correspondence between discrimination and goodness with finer detail. MDS is repeatedly shown to be effective at modeling perceptual similarity of vowels (Singh and Woods, 1971; Shepard, 1972; Fox, 1983; Kewley-Port and Atal, 1989). Various measures can be utilized to estimate raw similarity, including similarity judgments of sound pairs on a numerical scale, identification-confusion percentage (more frequent confusions for similar vowels), and reaction time measures (longer reaction times for more similar sounds). MDS assigns stimuli to a geometric space where distances in the Euclidean configuration correspond to perceived similarity (Borg and Groenen, 1997). In the MDS configuration, stimuli are constructed along major axes (dimensions) that are perpendicular to each other. Stimuli perceived as similar are close to each other in MDS space and stimuli perceived as dissimilar are further apart. Modeling similarity using MDS analysis uncovers relationships among stimuli that are difficult to observe solely from raw similarity ratings.

To support the prediction that restricted vowel working space area would affect speech intelligibility, the vowels of the largest vowel triangle were expected to be judged as more intelligible and rated as having higher ''goodness'' than those of the smallest vowel triangle. Moreover, the MDS was expected to show that perceptual distances between vowels of smaller triangles would be smaller than those of larger vowel triangles in the perceptual organization of normal listeners.

A. Methods

1. Participants

Twenty-two undergraduates (ten women) at universities in Taiwan were recruited from a solicitation posted on the Internet. A language background questionnaire was administrated before the experiment to ensure that Mandarin Chinese was the native or the dominant language for all participants. Participants reported no history of communication disorders and received $15 for their participation in this 1-h experiment.

2. Stimuli

The stimuli were 18 vowel tokens (three corner vowels /i/, /a/, and /u/×six variants) created using HLSYN speech synthesizer 2.2 (1996) and sampled at 11 025 Hz. Vowel variants were synthesized by manipulating the values of the first and second formants; the third, fourth, and fifth formants remained constant for the same phonetic category. Figure 2 shows the spatial relationship of point vowel variants on $F1/F2$ space. The psychoacoustic distances between neighboring tokens were equalized at 30 mels in $F1/F2$ vowel working space. The mel scale is essentially linear at low frequencies and logarithmic at high frequencies, and corresponds well to excitation patterns of the ear's basilar membrane. Fant (1973) argued that the mel scale was appropriate for vowels because the minimum perceptible differences were similar for the first three formants when measured in mels. Thus, vowel variants among neighboring tokens were scaled in this common metric.

The $F1$ and $F2$ values of the most intelligible talker with cerebral palsy (subject: CP11, vowel intelligibility $=100\%$, word intelligibility$=88\%$) in experiment 1 were utilized as the starting point to generate vowel variants in 30-mel steps. Then, the $F1$ and $F2$ values were gradually adjusted to reduce the vowel working space areas. The smallest vowel triangle of the synthesized vowel ($85\,190\ \mathrm{Hz}^2$) was smaller than the smallest vowel triangle of a talker with the cerebral palsy in experiment 1 (subject: CP7, area $=111\,157\ \mathrm{Hz}^2$). The spatial relationship on the $F1/F2$ plane of these synthesized vowel triangles was similar to that of the vowel triangles of normal talkers and talkers with cerebral palsy shown in Fig. 1 of experiment 1.

Table II lists the $F1$ and $F2$ values of vowel variants in mels. For /i/, $F1$ values increased from 298 to 367 Hz and $F2$ values decreased from 2232 to 1954 Hz; for /a/, $F1$ decreased from 794 to 623 Hz and $F2$ decreased from 1355 to 1292 Hz; for /u/, $F1$ increased from 370 to 443 Hz and $F2$ increased from 910 to 1090 Hz. The $F3$ and $F4$ values were 3186 and 3755 Hz for the /i/ variants, respectively; 2633 and 3648 Hz for the /a/ variants; and 2449 and 3518 Hz for the /u/ variants. The $F0$ of all tokens was 120 Hz and each vowel was 450 ms long. Mandarin Chinese is a tonal language, meaning that changes of pitch contours sign different lexical meanings for syllables. The $F0$ of vowel tokens did not change over time, so that the lexical tone pattern was similar to the high-level tone in Mandarin Chinese. The stimuli were equalized in rms power and played to subjects at a comfortable level of approximately 70 dB SPL.

TABLE II. $F1$ and $F2$ values (in mels) of synthetic vowel variants in experiment 2. The psychoacoustic distances between tokens were equalized at 30 mels in $F1/F2$ vowel working space. $F1$ and $F2$ values of the most intelligible vowels in experiment 1 were utilized as the starting point to generate vowel variants.

Psychoacoustic distances (in mels) from the clearest vowel	/i/		/a/		/u/	
	$F1$	$F2$	$F1$	$F2$	$F1$	$F2$
0	376	1692	843	1236	454	934
30	391	1666	814	1228	469	960
60	406	1640	785	1220	484	985
90	421	1614	756	1212	499	1011
120	436	1588	727	1205	514	1037
150	451	1563	698	1197	529	1063

3. Apparatus

The stimuli were presented and controlled by an Intel Pentium III laptop computer (Dell Inspiron 4100). The vowels were played bilaterally to subjects using noise-reduction headphones (Bose QuietComfort Acoustic Noise Canceling headset) while subjects were sitting in a sound-treated room. Subjects' responses were entered and recorded using the computer that controlled the presentation of stimuli.

4. Procedure

a. Phonetic identification and goodness ratings. Twenty participants were asked to identify each token as /i/, /a/, or /u/. After identification, they were asked to rate each token on a scale from 1 (poor) to 7 (good) regarding its goodness as a member of that phonetic category. At the start of the experiment, participants completed a practice session of 18 trials with each of the tokens presented once in a random order. After the practice, participants completed an experimental session of 36 trials (2 blocks of 18 tokens) with the order of trials randomized within each block.

b. Phonetic similarity judgment. Participants heard all possible pairs of the 18 tokens in an AX discrimination task. On each trial, the participant heard two tokens separated by 550 ms and were asked to decide whether the vowels in each pair of stimuli were the same or different. Their response for each trial (same or different) and reaction time (RT) were recorded. RTs on different trials were utilized as similarity scores and submitted to the multidimensional scaling (MDS) solution to create a perceptual map where the distance between any two tokens corresponded to perceived similarity.

Prior to the test, participants completed a short practice session. There were 12 ''same'' trials and 24 ''different'' trials chosen randomly from the trials used in the experimental session. The practice trials were the same as in the experimental session except that participants received feedback on their responses. During the test stage, 378 test pairs ($=306$ different trials and 72 same trials) were presented with no feedback. The same trials consisted of four trials of each of the 18 tokens. The 306 different trials consisted of each possible pair of the 18 tokens and the probability of presentation order, i.e., A–B and B–A, was equal. Participants were allowed to take a short break after every 40 trials; the entire session took 30 to 40 min to complete.

B. Results and discussion

1. Phonetic identification and goodness ratings

Figure 2 displays the results of phonetic identification and goodness ratings. There is high intersubject agreement on goodness rating (Cronbach $\alpha=0.95$) and phonetic identification for each vowel token (average identification percentage$=96.3\%$). Tokens with formant frequencies similar to the most intelligible vowels in experiment 1 were rated as better exemplars of vowel categories. Moreover, goodness judgments varied with vowel area in the $F1/F2$ space. As shown in Fig. 2, tokens composing larger vowel working spaces were generally judged as the better vowel exemplars. The tokens comprising the largest vowel triangle were judged to be the best exemplars for /i/ and /a/, and second best for /u/. In contrast, tokens composing the smallest vowel triangle were judged to be the poorest exemplars for each vowel category. Goodness ratings varied significantly within each vowel at the significance level of 0.001 in one-way repeated measure analyses of variance (ANOVAs), /i/, $F(5,95)=39.22$; /a/, $F(5,95)=22.97$; /u/, $F(5,95)=13.47$. Although there are equal psychophysical differences between neighboring tokens, tokens close to typical vowels in the $F1/F2$ space receive similar goodness ratings. A *posthoc* LSD analysis at $p=0.05$ demonstrates that the goodness ratings of better exemplars did not differ significantly for the three best (of six) /i/ tokens, the four best (of six) /a/ tokens, and the two best (of six) /u/ tokens. The finding that vowels

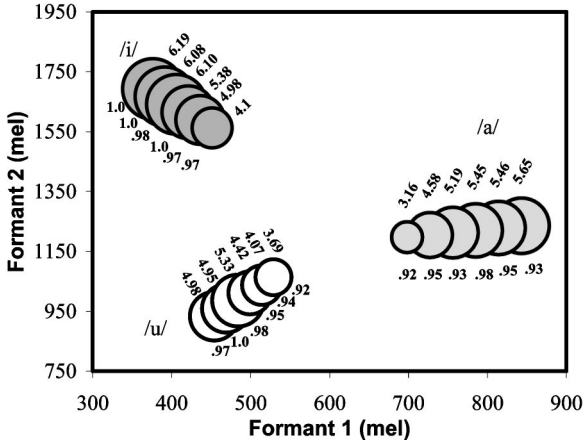


FIG. 2. Identification and goodness ratings for Mandarin vowels. For each token, the upper number lists goodness rating and the bottom number displays percent-correct identification. Circle size corresponds to perceived goodness.

composing larger acoustic space were judged to be better exemplars than vowels of smaller vowel triangles provides support for the prediction that vowel working space area greatly affects vowel intelligibility and goodness ratings.

Token goodness varies with acoustic vowel area. However, the shrunken acoustic space of synthesized tokens does not greatly affect the percentage of vowel identification (Fig. 2). The relatively stable identification percentages across tokens could reveal the categorical aspect of phonetic identification (e.g., Liberman *et al.*, 1961). However, the high identification percentage ($>90\%$) of tokens of the smallest vowel triangle could also be a result of the forced-choice identification task. The identification for tokens in the most restricted vowel space might be less consistent if participants were allowed to free-identify tokens as any one of the Mandarin vowels because their formant patterns were located at the adjacent area of different vowels on the $F1/F2$ space, e.g., the /i/ and /y/.

2. Similarity judgments and perceptual organization

Each subject's responses were put into the form of a lower triangular matrix composed of the log RT for each pair of tokens averaged across the presentation order. Responses to ''same'' trials were not considered in the analyses. Overall subject accuracy was 85.01%. Subjects correctly detected the stimulus pairs that were different on 87.2% of the trials and had false alarms on 17.2% of the trials of same sound pairs. Intersubject reliability was done to assess the consistency of RTs across all pairs of tokens. There was high agreement for response latencies on judging differences across all pairs (Cronbach $\alpha=0.98$).

The lower triangular matrix of log RTs was submitted to the ALSCAL (Takane, Young, and de Leeuw, 1977) MDS algorithm implemented by SPSS 10.0 (statistical software). The MDS solution used Kruskal's stress formula 1 (Kruskal, 1964), a Euclidean distance metric, and a monotonic regression function. This MDS solution placed the tokens in a two-dimensional space where the distances between tokens were fit into a monotonic function of the log RT. Preliminary MDS analyses demonstrated that the patterns of results was similar when RT and log RT were used, but analyses using log RT accounted for more variance. Furthermore, a previous study has successfully used log RT in the MDS analysis to examine the perceptual organization of vowels (Iverson and Kuhl, 1995). Figure 3 illustrates a two-dimensional MDS solution modeling the perceptual organization of point vowels with a stress of 0.241 ($R^2=0.801$).

The results of the MDS solution showed a larger perceptual triangle for vowels rated as good exemplars and a smaller triangle for vowels rated as poor exemplars, thus providing support for the prediction that reduced acoustic vowel working space results in shrunken perceptual space and adversely affects speech intelligibility. As shown in Fig. 3, the configuration of perceptual vowel organization is similar to the vowel acoustic triangle in the $F1/F2$ space. The perceptual configuration shown in Fig. 3 is consistent with a previous study that employed the MDS solution to examine vowel perceptual space (Singh and Woods, 1971). Moreover, better vowel exemplars composed larger perceptual triangles

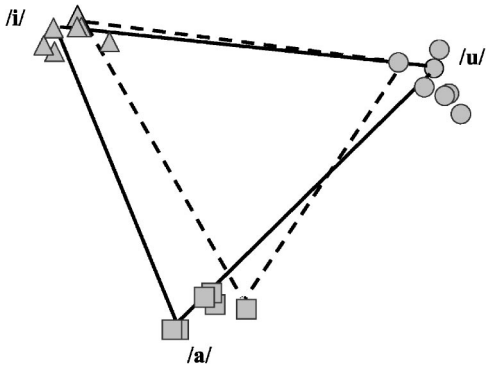


FIG. 3. A two-dimensional MDS solution shows the perceptual organization for perceiving Mandarin vowels. The solid lines connect the best vowel exemplars and dotted lines connect the poorest vowel exemplars.

than poor exemplars. The space formed by the best vowels (3.033 MDS distance^2) is 34% larger than that of the poorest vowels (2.2653 MDS distance^2). The shrunken perceptual space for vowels with reduced acoustic space further suggests that the reduced acoustic vowel working space of talkers with cerebral palsy decreases the intervowel perceptual distance and adversely affects speech intelligibility.

In addition to examining the effects of reduced acoustic vowel triangles on vowel perceptual space, this study explored within-vowel organization. As shown in the two-dimensional MDS solution of perceptual vowel working space (Fig. 3), perceptual clustering is more apparent for the best exemplars, but extended perceptual spacing occurs for the poorest exemplars (except for the poorest /i/). To further examine within-vowel organization, three lower triangular matrices of log RTs, one for each vowel, were analyzed separately in a one-dimensional MDS solution. The MDS solutions modeled the responses with a stress of 0.253 ($R^2=0.757$) for /i/, 0.236 ($R^2=0.790$) for /a/, and 0.239 ($R^2=0.771$) for /u/. Figure 4 displays these one-dimensional MDS solutions.

The relative positions of tokens in the one-dimensional MDS solutions (Fig. 4) corresponded similarly to their acoustic locations on the single vector, and, for within-vowel organization, better exemplars were clustered more tightly than the poorer exemplars for the three corner vowel categories. The perceptual clustering and spacing shown in Figs. 3 and 4 can be associated with the ''perceptual magnet effect (PME)'' (Grieser and Kuhl, 1989; Kuhl, 1991; Iverson and Kuhl, 1995, 1996; Iverson *et al.*, 2003; Miller and Eimas, 1996). The PME posits that phonetic prototypes (i.e., good exemplars) function as perceptual magnets in the perceptual space for phonetic categories; that is, perceptual distance is shrunken around good exemplars, demonstrating greater per-

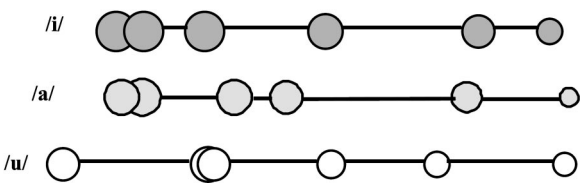


FIG. 4. One-dimensional MDS solutions show the internal organization of Mandarin corner vowels. Circle size corresponds to perceived goodness.

ceptual similarity; it is extended around poor exemplars, demonstrating less perceptual similarity. The PME might mediate the perceptual distortion shown in Fig. 4, which shows that acoustically equal spacing among vowel variants is unequally mapped onto the perceptual space.

IV. GENERAL DISCUSSION

The goal of this study was to examine whether the restriction of articulatory movement is associated with the intelligibility of speech produced by Mandarin-speaking young adults with cerebral palsy. The results of acoustic analysis indicate that talkers with cerebral palsy produced a smaller vowel working space area than normal talkers. The reduced vowel working space area reflects the underlying speech problem associated with the speech intelligibility deficit. That is, the more restricted tongue/jaw elevation and anterior–posterior movements contract the vowel working space area. Moreover, for talkers with cerebral palsy, correlational analysis demonstrates a significant association between the vowel working space areas and speech intelligibility, further suggesting that reduced vowel working space area is strongly associated with listeners' difficulty in correctly identifying vowels.

What is the perceptual effect of a reduced vowel working space on listeners' perceptual organization? In experiment 2, we systematically manipulatied vowel working space area. The results of the speech perception experiment indicate that reduced acoustic vowel working space is associated with a reduced perceptual space. Furthermore, the results of the MDS analyses reveal that the category goodness of vowel tokens affects the nonlinear transformation between acoustic space and perceptual space. When examining within-vowel perceptual organization, perceptual clustering is apparent for better vowel exemplars but not poorer exemplars.

A. Relationship between vowel working space area and speech intelligibility

The present study demonstrates a strong relationship between vowel acoustic space area encompassed by the three corner vowels, and speech intelligibility measured with words produced by Mandarin talkers with cerebral palsy. Therefore, vowel working space area can serve as a general indicator of speech intelligibility. The likelihood of reduced vowel space probably depends on the severity of dysarthria. Measures of vowel working space area have rarely, if ever, been reported for talkers with cerebral palsy. The present study increases our understanding of specific aspects of decreased speech intelligibility in this group of talkers. Larger vowel working spaces were reported to correlate with better speech intelligibility in some talkers without speech disorders (Bradlow *et al.*, 1996), and, some studies have shown reduced vowel working space in English talkers with various types of dysarthria (Turner *et al.*, 1995; Weismer *et al.*, 2001, Ziegler and von Cramon, 1983). The results of the present study are consistent with previous findings on English talkers with similar diagnoses of motor speech impairment, suggesting that the extent of vowel working space could be helpful for representing the level of speech intelligibility across languages. Similar findings among English and Mandarin talkers with dysarthria support the argument that vowel working space has predictive power for speech intelligibility. Moreover, this could be independent of both language and type of dysarthria.

However, in addition to vowel formant distortion, additional speech features that are common in the dysarthric speech secondary to cerebral palsy, such as breathy voice quality, varied vocal intensity, and unstable speaking rate, could also reduce the speech intelligibility of atypical talkers (e.g., Kent *et al.*, 1978). It is possible that an individual's speaking rate might be associated with format variations in experiment 1. Speaking rate was shown to be related to the perceived intelligibility of talkers with dysarthria (Turner *et al.*, 1995). For example, ALS talkers' sentences were less intelligible when using their fast speaking rate compared to their normal rate. Could speaking rate mediate the relationship between vowel working space area and speech intelligibility shown in this study? When typical talkers and talkers with dysarthria secondary to multiple sclerosis and Parkinson's disease used various speaking rates to read a passage, vowel working space area for their slow speaking rate was significantly larger than that of their normal or fast speaking rate (Turner *et al.*, 1995; Tjaden and Wilding, 2004). In addition, infant-directed speech is characterized by a slower speaking rate and exaggerated pitch variation when compared with adult-directed speech (Fernald and Simon, 1984; Kuhl *et al.*, 1997; Liu *et al.*, 2003). Mothers of different native languages consistently utilize a larger vowel working space area when addressing their infants as opposed to adults (Kuhl *et al.*, 1997; Liu *et al.*, 2003). For typical talkers, however, even though infant-directed speech exhibits larger vowel working space areas than adult-directed speech, an individual's speaking rate, as measured in vowel duration, is not associated with that individual's vowel working space area (Liu, 2002). Results of this study cannot determine whether the individual's speaking rate mediates the association between vowel working space and speech intelligibility because this study did not assess the relative contribution of speaking rate and vowel space to speech intelligibility.

B. Listeners' perceptual organization for reduced vowel working space

Perceptual characteristics of vowels from a reduced vowel working space are further examined in experiment 2, where the vowel working space areas were systematically reduced through manipulation of $F1$ and $F2$. As predicted from the acoustic analysis on the vowel working space, typical listeners rated tokens from smaller vowel working spaces as poorer exemplars of vowels. In contrast, listeners judged tokens from larger vowel working spaces as better exemplars. Using token-pair similarity judgments, reaction time, and multidimensional scaling to examine the effect of restricted vowel working space on perceptual organization, the poor vowel exemplars or less intelligible tokens were represented with smaller perceptual space areas than good exemplars or more intelligible tokens. In other words, the intervowel perceptual distances of less intelligible vowels were reduced compared to those of more intelligible vowels.

In addition to revealing the perceptual constraints of perceiving vowels of reduced space, the results of within-vowel perceptual organization showed that the perceptual distance among vowel tokens is reduced near better exemplars but expanded near poor exemplars. This perceptual organization could be due to the perceptual magnet effect (e.g., Kuhl, 1991). Although there is a debate about the source of perceptual distortion and the stability of phonetic prototypes for speech sounds (Frieda *et al.*, 1999; Lotto, Kluender, and Holt, 1998), the results of the current data imply that the perceptual magnet effect might play a role in the perceptual organization of speech sounds. The vowels of talkers with cerebral palsy are judged as poor exemplars or as less intelligible, and are represented in perceptual space farther away from the better exemplars or more intelligible vowels that normal listeners generally encounter and can easily process. Therefore, vowels of talkers with cerebral palsy are not perceptually close to intelligible vowels. Vowels produced by atypical talkers are likely to require listeners to allocate more perceptual resources. This processing inefficiency increases the difficulty when listeners attempt to identify the vowels of cerebral-palsied speakers.

Combining the results of both the acoustic and perceptual experiments, the distorted vowels of talkers with cerebral palsy compose a smaller acoustic space.. The results of our perceptual organization tests demonstrate that vowels from talkers with cerebral palsy are represented with reduced intervowel perceptual distance compared to vowels from typical talkers. This increases the difficulty normal listeners have in mapping the degraded acoustic signals onto existing phonetic categories.

Results of this study provide segmental evidence for reduced intelligibility from both talkers' and listeners' points of view. In addition to the acoustic characteristics of dysarthric speech, listeners' experience helps them to decode the phonetic and lexical information of degraded speech, and this improves speech intelligibility (e.g., Liss *et al.*, 2002). Listening experience can alter the perceptual organization of speech sounds in neurologically intact people. For example, when presented with the same set of stimuli varying in the $F2$ and $F3$ of /r/ and /l/ consonants, the perceptual organization of American English speakers is extended along the $F3$ dimension, the critical acoustic cue distinguishing /r/ from /l/, but Japanese-speaking listeners' perceptual organization was extended along $F2$ (Iverson and Kuhl, 1996; Iverson *et al.*, 2003). Studies directed towards the relationship between familiarity with degraded acoustical signals and perceptual organization in future studies will not only contribute to theory but also to the clinical goal of designing more efficient programs to help listeners better understand speech produced by talkers with cerebral palsy.

ACKNOWLEDGMENT

The authors would like to thank Tobey Nelson for her helpful comments on an earlier draft of the manuscript. The efforts of reviewers, especially Gary Weismer, are deeply appreciated.

Ansel, B. M., and Kent, R. D. (**1992**). ''Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy,'' J. Speech Hear. Res. **35**, 296–308.

Borg, I., and Groenen, P. (**1997**). *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York).

Bradlow, A. R., Torretta, G. M., and Posoni, D. B. (**1996**). ''Intelligibility of normal speech. I. Global and fine-rained acoustic-phonetic talker characteristics,'' Speech Commun. **20**, 255–272.

Bunton, K., and Weismer, G. (**2001**). ''The relationship between perception and acoustics for a high–low vowel contrast produced by talkers with dysarthria,'' J. Speech Hear. Res. **44**, 1215–1228.

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (**2002**). ''What's new, pussycat? On talking to babies and animals,'' Science **296**, 1435–1435.

Eimas, P. D., and Miller, J. L. (**1992**). ''Organization in the perception of speech by young infants,'' Psychol. Sci. **3**, 340–345.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT Press, Cambridge, MA).

Fernald, A., and Simon, T. (**1984**). ''Expanded intonation contours in mothers' speech to newborns,'' Dev. Psychol. **20**, 104–113.

Fox, R. A. (**1983**). ''Perceptual structure of monophthongs and diphthongs in English,'' Lang. Speech **26**, 21–60.

Frieda, E. M., Walley, A. C., Flege, J. E., and Sloane, M. E. (**1999**). ''Adults' perception of native and nonnative vowels: Implication for the perceptual magnet effect,'' Percept. Psychophys. **61**, 561–577.

Grieser, D. L., and Kuhl, P. K. (**1989**). ''Categorization of speech by infants: Support for speech-sound prototypes,'' Dev. Psychol. **25**, 577–588.

Hardy, J. C. (**1983**). *Cerebral Palsy* (Prentice-Hall, Englewood Cliffs, NJ).

Higgins, C. M., and Hodge, N. M. (**2001**). ''$F2/F1$ vowel quadrilateral area in young children with and without dysarthria,'' Can. Acoust. **29**, 66–68.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

HLSYN (**1996**). Computer software (Sensimetrics Corporation, Cambridge, MA).

Hodge, M. M. (**1999**). ''Relationship between $F2/F1$ vowel quadrilateral area and speech intelligibility in a child with progressive dysarthria,'' Can. Acoust. **27**, 84–85.

Howie, J. M. (1976). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge).

Iverson, P., and Kuhl, P. K. (**1995**). ''Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling,'' J. Acoust. Soc. Am. **97**, 553–562.

Iverson, P., and Kuhl, P. K. (**1996**). ''Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/,'' J. Acoust. Soc. Am. **99**, 1130–1140.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. and Siebert C. (**2003**). ''A perceptual interference account of acquisition difficulties for non-native phonemes,'' Cognition **87**, B47–57.

Kent, R. D., Dembowski, J., and Lass, N. J. (**1995**). ''The acoustic characteristics of American English,'' in *Principles of Experimental Phonetics*, edited by N. J. Lass (Masby-Year Book, St. Louis, MO), pp. 185–225.

Kent, R. D., Kent, J. F., Rosenbek, J. C., Voperian, H. K., and Weismer, G. (**1997**). ''A speaking task analysis of the dysarthria in cerebellar disease,'' Folia Phoniatr. Logop. **49**, 63–82.

Kent, R. D., Netsell, R., and Abbs, J. H. (**1978**). ''Articulatory abnormalities in athetoid cerebral palsy,'' J. Speech Hear. Disord. **43**, 353–373.

Kent, R. D., Weismer, G., Kent, J. F., and Rosenbek, J. C. (**1989**). ''Toward explanatory intelligibility testing in dysarthria,'' J. Speech Hear. Disord. **54**, 482–499.

Kewley-Port, D., and Atal, B. S. (**1989**). ''Perceptual differences between vowels located in a limited phonetic space,'' J. Acoust. Soc. Am. **85**, 1726–1740.

Krause, J. C., and Braida, L. D. (**2004**). ''Acoustic properties ofnaturally produced clear speech at normal speaking rates,'' J. Acoust. Soc. Am. **115**, 362–378.

Kruskal, J. B. (**1964**). ''Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis,'' Psychometrika **29**, 1–27.

Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for prototypes of speech categories, monkeys do not,'' Percept. Psychophys. **50**, 93–107.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda F. (**1997**). ''Cross-language analysis of phonetic units in language addressed to infants,'' Science **277**, 684–686.

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Malden, MA).

Liberman, A. M., Harris, K. S., Kinney, J. A., and Lane, H. L. (**1961**). ''The discrimination of relative-onset time of the components of certain speech and nonspeech patterns,'' J. Exp. Psychol. **61**, 379–388.
Lindblom, B. (**1990**). ''Explaining phonetic variation: A sketch of H and H theory,'' in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht, The Netherlands), pp. 403–439.
Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (**2002**). ''The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria.'' J. Acoust. Soc. Am. **112**, 3022–3030.
Liu, H.-M. (**2002**). ''The acoustic-phonetic characteristics of infant-directed speech in Mandarin Chinese and their relation to infant speech perception in the first year of life,'' unpublished Ph.D. dissertation, University of Washington.
Liu, H. M., Kuhl, P. K., and Tsao, F. M. (**2003**). ''The association between mothers' speech clarity and infants' speech discrimination skill,'' Dev. Sci. **6**, F1–F10.
Liu, H. M., Tseng, C. H., and Tsao, F. M. (**2000**). ''Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy,'' Clin. Linguist. Phon. **14**, 447–464.
Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1998**). ''Depolarization the perceptual magnet effect,'' J. Acoust. Soc. Am. **96**, 2076–2087.
McDonald, E. T. (**1987**). ''Cerebral palsy: Its nature, pathogenesis, and management,'' in *Treating Cerebral Palsy*, edited by E T. McDonald (Pro-Ed, Austin, TX), pp. 1–19.
Miller, J. L., and Eimas, P. D. (**1996**). ''Internal structure of voicing categories in early infancy,'' Percept. Psychophys. **58**, 1157–1167.
Peterson, G., and Barney, H. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Platt, L. J., Andrews, G., and Howie, P. M. (**1980**). ''Dysarthria of adult cerebral palsy. II. Phonemic analysis of articulation errors,'' J. Speech Hear. Res. **23**, 41–55.
Platt, L. J., Andrews, G., Young, M., and Quinn, P. T. (**1980**). ''Dysarthria of adult cerebral palsy. I. Intelligibility and articulatory impairment,'' J. Speech Hear. Res. **23**, 28–40.
Remez, R. E. (**2001**). ''The interplay of phonology and perception considered from the perceptive of perceptual organization,'' in *The Role of Speech Perception in Phonology*, edited by E. Hume and K. Johnson (Academic, San Diego, CA), pp. 27–52.
Shepard, R. (**1972**). ''Psychological representation of speech sounds,'' in *Human Communication: A Unified View*, edited by E. E. David and P. B. Denes (McGraw Hill, New York), pp. 67–113.
Singh, S., and Woods, D. R. (**1971**). ''Perceptual structure of 12 American vowels,'' J. Acoust. Soc. Am. **49**, 1861–1866.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Takane, Y., Young, F. W., and De Leeuw, J. (**1977**). ''Nonmetric individual differences multidimensional scaling: An alternating least-squares method with optimal scaling features,'' Psychometrika **42**, 7–67.
Tjaden, K., and Wilding, G. E. (**2004**). ''Rate and loudness manipulations in dysarthria: Acoustic and perceptual fundings,'' J. Speech Lang. Hear. Res. **47**(4), 766–783.
Tseng, C.-Y. (**1990**). *An Acoustic Phonetic Study on Tones in Mandarin Chinese* (Institute of History & Philology Academia Sinica, Taipei, Taiwan).
Turner, G. S., Tjaden, K., and Weismer, G. (**1995**). ''The influence of speaking rate of vowel working space and speech intelligibility for individuals with amyotrophic lateral sclerosis,'' J. Speech Hear. Res. **38**, 1001–1013.
Weismer, G., and Martin, R. E. (**1992**). ''Acoustic and perceptual approaches to the study of intelligibility,'' in *Intelligibility in Speech Disorders: Theory, Measurement, and Management*, edited by R. D. Kent (John Benjamins, Amsterdam, The Netherlands), pp. 67–118.
Weismer, G., Martin, R. E., Kent, R. D., and Kent, J. F. (**1992**). ''Formant trajectory characteristics of males with amyotrophic lateral sclerosis,'' J. Acoust. Soc. Am. **91**, 1085–1098.
Weismer, G., Jeng, J-Y., Laures, J. S., Kent, R. D., and Kent, J. F. (**2001**). ''Acoustic and intelligibility characteristics of sentence production in neuorgenic speech disorders,'' Folia Phoniatr. Logop. **53**, 1–18.
Weismer, G., Laures, J. S., Jeng, J-Y., Kent, R. D., and Kent, J. F. (**2000**). ''Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis.'' Folia Phoniatr. Logop. **52**, 201–219.
Whitehill, T. L., and Ciocca, V. (**2000**). ''Perceptual-phonetic predictors of single-word intelligibility: A study of Cantonese dysarthria,'' J. Speech Hear. Res. **43**, 1451–1465.
Wu, J. T., and Liu, I. M. (**1987**). ''A study on the sound properties and semantic classifications of Chinese words,'' Research report of National Science Council, Republic of China (Taiwan).
Yorkston, K. M., Beukelman, D. R., and Bell, K. R. (**1988**). *Clinical Management of Dysarthric Talkers*. (Little, Brown, Boston).
Yorkston, K., Dowden, P. A., and Beukelman, D. R. (**1992**). ''Intelligibility measurement as a tool in the clinical management of dysarthric talkers,'' in *Intelligibility in Speech Disorders: Theory, Measurement, and Management*, edited by R. D. Kent (John Benjamins, Amsterdam, The Netherlands), pp. 265–285.
Ziegler, W., and von Cramon, D. (**1983**). ''Vowel distortion in traumatic dysarthria: A formant study,'' Phonetica **40**, 63–78.
Ziegler, W., and von Cramon, D. (**1986**). ''Spastic dysarthria after acquired brain injury: An acoustic study,'' Br. J. Disord. Commun. **21**, 173–188.

Perception of familiar contrasts in unfamiliar positions[a)]

Mirjam Broersma[b)]
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 24 August 2004; revised 21 March 2005; accepted 22 March 2005)

This paper investigates the perception of non-native phoneme contrasts which exist in the native language, but not in the position tested. Like English, Dutch contrasts voiced and voiceless obstruents. Unlike English, Dutch allows only voiceless obstruents in word-final position. Dutch and English listeners' accuracy on English final voicing contrasts and their use of preceding vowel duration as a voicing cue were tested. The phonetic structure of Dutch should provide the necessary experience for a native-like use of this cue. Experiment 1 showed that Dutch listeners categorized English final /z/–/s/, /v/–/f/, /b/–/p/, and /d/–/t/ contrasts in nonwords as accurately as initial contrasts, and as accurately as English listeners did, even when release bursts were removed. In experiment 2, English listeners used vowel duration as a cue for one final contrast, although it was uninformative and sometimes mismatched other voicing characteristics, whereas Dutch listeners did not. Although it should be relatively easy for them, Dutch listeners did not use vowel duration. Nevertheless, they attained native-like accuracy, and sometimes even outperformed the native listeners who were liable to be misled by uninformative vowel duration information. Thus, native-like use of cues for non-native but familiar contrasts in unfamiliar positions may hardly ever be attained. [DOI: 10.1121/1.1906060]

PACS numbers: 43.71.Hw, 43.71.Es [ARB] Pages: 3890–3901

I. INTRODUCTION

In 1939, Trubetzkoy (reprinted as Trubetzkoy, 1977) observed that the sounds of a foreign language often get misinterpreted, because they go through the ''phonological sieve'' of the native language. Later research has proven Trubetzkoy right. The Perceptual Assimilation Model (PAM) (Best, 1994; Best, McRoberts, and Sithole, 1988) describes how listeners assimilate non-native speech sounds to the native category that is perceptually most similar. The PAM predicts which non-native speech sounds will be difficult to distinguish, based on the similarities and dissimilarities of the phonological structures of the native and the non-native language. The most difficult distinction is that between non-native speech sounds which match a single native category equally well. If the non-native language has two categories where the native language has only one in the same phonetic space, both non-native speech sounds will be assimilated to a single category. This is the case, for example, with Japanese listeners' perception of English /r/ and /l/ (Best and Strange, 1992). According to the PAM the easiest distinction is that between non-native speech sounds which are assimilated to two separate native categories. As the non-native contrast corresponds to a native contrast, it is easy to perceive.

However, languages not only have a phoneme inventory, they also have their own language-specific phonotactic constraints. The perception of non-native contrasts not only depends on the presence or absence of similar speech sounds in the native language, but also on native-language phonotactic constraints. This was demonstrated in a study of Chinese listeners' perception of the English /d/–/t/ contrast in word-final position (Flege, 1989). Chinese has a /d/–/t/ contrast, but not in word-final position. Word-initial /d/ and /t/ are not distinguished by closure voicing in Chinese, but on the basis of information in the release burst (Flege, 1989). Flege (1989) found that Chinese learners of English categorized unedited tokens of English word-final /d/ and /t/ almost as accurately as the native English listeners did. Their performance hardly decreased when closure voicing was removed, but was strongly affected by removal of the release burst. Flege concluded that the Chinese listeners used Chinese word-initial cues to distinguish between English /d/ and /t/ in word-final position. Flege and Wang (1989) showed that not only experience with the contrast itself, but native-language experience with any word-final stops influenced the perception of the word-final stop voicing contrast. Neither Cantonese Chinese nor Mandarin Chinese has a word-final stop voicing contrast, but /p,t,k/ can occur word finally in Cantonese, whereas Mandarin does not permit any word-final obstruents. Flege and Wang (1989) found that native listeners of Cantonese distinguished the English final /d/–/t/ contrast more accurately than native listeners of Mandarin did, which they attributed to the Mandarin listeners' lack of native language experience with word-final obstruents.

Further, the perception of non-native contrasts may depend not only on the presence or absence of similar phonemes in the native language, but also on the presence of utterly different contrasts. Crowther and Mann (1992) showed that the use of perceptual cues for a particular non-native contrast may depend on the use of the same cues for other phoneme distinctions in the native language. Like Mandarin Chinese, Japanese has a /d/–/t/ contrast, and does not permit word-final stops. Whereas Japanese has long and short vowels, Mandarin Chinese does not have this distinction. Crowther and Mann tested the perception of the English

[a)]A partial report of this work was presented at the 147th Meeting of the Acoustical Society of America, New York, May 2004.
[b)]Electronic mail: mirjam.broersma@mpi.nl

word-final /d/–/t/ contrast by Japanese and Mandarin learners of English. The Japanese listeners showed a greater sensitivity to the duration of the vowel preceding the final consonant and categorized the English final /d/–/t/ contrast more accurately than the Mandarin listeners did.

Thus, the categorization of a non-native contrast which exists in the native language, but in a position where it does not occur in the native language, seems to benefit from native-language experience with one of the phonemes of the contrast in the relevant position and from experience with relevant perceptual cues. These findings suggest that the potential for accurate and native-like categorization of a non-native but familiar contrast in an unfamiliar position is highest for native listeners of a language which provides such experience. Of all languages that contrast voiced and voiceless obstruents but not in word-final position, those languages which allow for either voiced or voiceless obstruents in word-final position, and in which vowel duration is used as a cue (for any phoneme contrast), offer the best preparation for accurate categorization of the word-final obstruent voicing contrast and for the use of vowel duration as a cue. As Dutch has a distinction between voiced and voiceless obstruents in word-initial and -medial position, allows for voiceless obstruents in word-final position, and also provides experience with the use of vowel duration as a cue for several phoneme distinctions, native listeners of Dutch should be well prepared to learn to distinguish English voiced and voiceless word-final obstruents as a familiar contrast in an unfamiliar position, and to use vowel duration as a cue. Especially advanced learners of English can be expected to have learned to do this, through combining their native and non-native language experience. Therefore, this paper investigates whether Dutch listeners with a high level of proficiency in English categorize English final obstruent voicing contrasts with a native-like level of accuracy and with a native-like use of the vowel duration cue. It provides a test of the perception of a non-native but familiar contrast in an unfamiliar position by listeners with a language background that is most suitable for the task.

Dutch and English share four pairs of voiced and voiceless obstruents: the alveolar and labiodental fricatives /z/, /s/, /v/, and /f/, and the bilabial and alveolar stops /b/, /p/, /d/, and /t/. Unlike English, Dutch neutralizes voicing distinctions in syllable-final, prepausal position (Booij, 1995). Thus, although in Dutch obstruent voicing is a relevant contrast in word-initial and-medial position, Dutch has no word-final voicing contrasts. Dutch does allow for /s,f,p,t/ in word-final position. Further, Dutch distinguishes between long and short vowels (Booij, 1995). As part of the difference between long and short vowels is phonetic vowel duration, Dutch listeners are familiar with the assessment of this cue. Dutch listeners even have native-language experience with the use of vowel duration as a cue to word-medial obstruent voicing. In Dutch, vowels preceding a medial voiced consonant are slightly longer than vowels preceding a medial voiceless consonant. According to Slis and Cohen (1969a), the average difference is 30 ms before stops and 40 ms before fricatives. Van den Berg (1989) found that Dutch listeners used vowel duration to decide on the voicing of intervocalic two-obstruent sequences, although it was not among the most important perceptual cues. A study by Jongman *et al.* (1992) suggests that Dutch listeners may be able to generalize their knowledge about the relationship between vowel duration and word-medial obstruent voicing to the case of word-final obstruents. In this study, Dutch listeners categorized vowels from a vowel length continuum as long or short. Stimuli corresponded to the Dutch words /stad/ and /sta:t/, and /zat/ and /za:d/, in which vowel length and underlying voicing of the final consonant are crossed. The surface word-final consonant was always voiceless. The location of the phoneme boundary differed between the two continua, suggesting that the perception of ambiguous vowel duration depended on the underlying voicing of the word-final stop.

In English, the difference in vowel duration before voiced and voiceless obstruents is larger than in Dutch. Peterson and Lehiste (1960) found a difference of 96 ms before word-final stops and 148 ms before word-final fricatives. There is extensive evidence for the great importance of preceding vowel duration for the perception of voicing of word-final obstruents in English (e.g., Raphael, 1972). Although the role of vowel duration as a cue to voicing seems to be smaller in Dutch than in English, Dutch listeners' familiarity with the cue in word-medial position may facilitate its use in word-final position in English.

As their native language has not provided them with any knowledge about the relevant acoustic cues for voicing in final position, Dutch listeners may try to identify the voicing of English final obstruents with the aid of the perceptual cues they rely on for Dutch initial and intervocalic voicing contrasts. This may be quite successful, as Dutch and English obstruents have a high degree of articulatory similarity, and the perceptual cues that signal the voicing distinctions overlap to some extent. Van Alphen and Smits (2004) showed that Voice Onset Time (VOT), specifically the presence or absence of prevoicing, is the strongest cue to initial stop voicing in Dutch. In the absence of prevoicing, voicing judgments for labials relied most strongly on the extent of $F0$ change into the following vowel, and for alveolars on the spectral center of gravity of the burst. Other significant cues were the duration and power of the burst. For intervocalic obstruents, presence or absence of vocal-fold vibration (Slis and Cohen, 1969b; Slis and Van Heugten, 1989), closure duration for stops (Kuijpers, 1996; Slis and Cohen, 1969a), and frication duration for fricatives (Slis and Van Heugten, 1989) have been shown to influence the perception of voicing. For intervocalic two-obstruent sequences, presence or absence of vocal-fold vibration during the closure of the two obstruents is the most important cue (Van den Berg, 1989). Closure duration of the second consonant, duration of the preceding vowel, and for fricatives the intensity of frication noise play a smaller but significant role (Van den Berg, 1989).

All of the above-mentioned cues have been found to be used by English listeners to distinguish voiced and voiceless obstruents in word-final position (see, e.g., Watson, 1983 for a review), and Dutch listeners may use their knowledge about Dutch voiced and voiceless obstruents to make the same distinction for final English obstruents. However, there

are differences between Dutch and English obstruents, and thus between the critical values of the cues, and the weight attributed to each cue for optimal identification.

For example, the two languages differ in the critical value of VOT for voicing of initial stops. English contrasts voiceless unaspirated and voiceless aspirated stops (Lisker and Abramson, 1964), and the duration of the voicing lag is a cue to voicing in English (e.g., Watson, 1983). Van Alphen and Smits (2004) found that 75% of Dutch voiced initial stops were produced with a voicing lead, and that the presence or absence of prevoicing was the strongest perceptual cue for initial stop consonant voicing in Dutch. Although initial stops without prevoicing were not automatically categorized as voiceless, but were assessed on the basis of other cues (as described above), initial stops without prevoicing were misperceived more often than prevoiced stops (37% vs 1%). In English, initial voiced stops are less often prevoiced than in Dutch. Smith (1978) found that bilabial voiced stops were prevoiced 56%, and alveolar stops 50% of the time in careful speech. Therefore, Dutch listeners may misperceive English initial voiced stops relatively often.

Another difference between Dutch and English is the importance of the duration of the preceding vowel as a cue to obstruent voicing. If Dutch listeners process English final obstruents in the same way they process Dutch obstruents, they may not attribute as much weight to vowel duration as English listeners do. This may not be a problem when enough other cues are available, but it may lead to less accurate categorization of unreleased stops. In English, final stops are often produced without a release burst (Byrd, 1993). English listeners have little difficulty identifying the voicing of stops without a release burst (e.g., Flege and Hillenbrand, 1987), which may be explained by the redundancy of information in the speech signal. However, if Dutch listeners use vowel duration as a voicing cue less than English listeners do, the Dutch listeners may have more difficulty identifying English final obstruents without a release burst.

In experiment 1, Dutch and English listeners' categorization of the British English obstruent voicing contrasts /z/–/s/, /v/–/f/, /b/–/p/, and /d/–/t/ was investigated in initial and final position in nonwords. It was investigated whether Dutch listeners had a preference for identifying English final obstruents as voiceless, as Dutch allows voiceless but not voiced obstruents in word-final position. The effect of removal of the release burst was investigated for the final stops. For reasons of comparison, a contrast which was expected to be difficult to distinguish for Dutch listeners was included in the experiment, namely the English /æ/–/ɛ/ contrast. The PAM predicts that this phoneme pair belongs to the set of most difficult English contrasts for Dutch listeners. Standard southern British English distinguishes two open midfront unrounded vowels, whereas Dutch has only one vowel in this part of the vowel space. Although the Dutch vowel is denoted as /ɛ/, it is lower than the English /ɛ/, so that it is located between English /ɛ/ and /æ/. As Dutch listeners will assimilate both English vowels to the single Dutch category, the distinction between the phonemes is expected to be difficult. Indeed, British English /æ/ and /ɛ/ have been found to be difficult to distinguish for Dutch listeners (Schouten, 1975).

TABLE I. Experiment 1 items.

/æ/–/ɛ/	/fæf/–/fɛf/
Initial /z/–/s/	/**z**i:f/–/**s**i:f/
Final /z/–/s/	/fu:**z**/–/fu:**s**/
Initial /v/–/f/	/**v**u:k/–/**f**u:k/
Final /v/–/f/	/ku:**v**/–/ku:**f**/
Initial /b/–/p/	/**b**o:f/–/**p**o:f/
Final /b/–/p/	/fi:**b**/–/fi:**p**/
Initial /d/–/t/	/**d**i:s/–/**t**i:s/
Final /d/–/t/	/fo:**d**/–/fo:**t**/

Dutch and English listeners' use of vowel duration as a cue to final obstruent voicing was further investigated in experiment 2. If Dutch listeners use vowel duration as a cue less than English listeners do, Dutch listeners may find it easier to ignore vowel duration when this cue is made unreliable than English listeners do. In experiment 2 it was investigated whether Dutch and English listeners relied on vowel duration as a cue to final obstruent voicing when this cue was uninformative and when it mismatched with other information in the signal.

II. EXPERIMENT 1

A. Method

1. Participants

Twenty native speakers of Dutch and 20 native speakers of British English took part in the experiment. The Dutch participants had a high level of proficiency in English as a second language. They had received on average 7 years of English instruction in primary and secondary education. The English participants did not know any Dutch. The Dutch participants were recruited from the Max Planck Institute participant pool, and the English participants from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex. None reported any hearing loss. All were volunteers and received a small fee for participation.

2. Materials

The vowel contrast /æ/–/ɛ/ was to be tested in one position and the four consonant contrasts /z/–/s/, /v/–/f/, /b/–/p/, and /d/–/t/ in two positions. Therefore, nine pairs of monosyllabic CVC items were selected. Each pair differed in one phoneme pair, corresponding to the contrast to be tested. The nontarget consonants in the CVC items were obstruents, in order to minimize their influence on the target sounds. All

TABLE II. Experiment 1, acoustic measures of stimuli with /æ/ and /ɛ/: Mean $F1$ steady-state frequency (Hz), mean $F2$ steady-state frequency (Hz), and mean vowel duration (ms).

	/æ/	/ɛ/
$F1$	824	744
$F2$	1602	1823
Vowel duration	167.4	131.6

TABLE III. Experiment 1, acoustic measures of stimuli with initial and final fricatives: Mean vowel duration (ms), mean fricative duration (ms), and mean fricative power above 500 Hz (logarithm of the spectral power of the frication noise above 500 Hz in Pa^2).

	Initial				Final			
	/z/	/s/	/v/	/f/	/z/	/s/	/v/	/f/
Vowel duration	⋯	⋯	⋯	⋯	258.7	130.5	264.9	118.5
Fricative duration	112.8	178.3	111.2	160.9	160.5	257.8	144.8	232.7
Fricative power	−2.2	−2.2	−3.1	−3.1	−2.6	−2.2	−3.6	−3.3

items were nonwords in Dutch and English, according to the CELEX database (Baayen, Piepenbrock, and Gulikers, 1995). The items are presented in Table I.

The materials were recorded by a male native speaker of British English. The speaker read the items one by one, separated by a pause, in a clear citation style. The recording was made in a soundproof booth with a Sennheiser microphone and stored directly onto a computer at a sample rate of 16 kHz. For each target phoneme in each position, three tokens were extracted from the file with the speech editor PRAAT. For the items with a final stop, only tokens with a clearly audible release burst were selected. These tokens were kept unedited for the condition with release burst, and for the condition without release burst the signal was truncated at the last positive zero crossing before the release burst.

Acoustic measurements were made of several characteristics which may be relevant for the distinction of the contrasts. The results are presented in Table II for the target vowels, in Table III for the fricatives, and in Table IV for the stops.

3. Design

Each fricative contrast occurred in initial and final position. The stops occurred in three conditions: initial position, final position with release burst, and final position without release burst. The order of presentation of the initial and final positions was counterbalanced. As the items in the two final conditions were based on the same tokens, the final with release burst condition always occurred after the final without release burst condition. The target phonemes /s/ and /f/ also occurred as nontargets in stimuli for other contrasts. The blocks were ordered such that the subjects were not exposed to a phoneme before the contrast it was part of was being tested.

The items were presented in 11 blocks, each block representing one phoneme contrast in one condition. Each block consisted of four repetitions of six tokens, semirandomized such that the same phoneme occurred maximally five times in succession and the same token maximally once.

4. Procedure

Participants were tested one at a time in a quiet room. They were informed in their native language that they would hear a series of nonwords, which would be similar except for one sound. They were instructed to decide which one of two alternatives this sound was, and to indicate their response with a button press. Before each block, they received further information about the two response alternatives in that block, and about the position of the target phoneme. They were not instructed about the truncation in the condition without release burst. Before the /æ/–/ɛ/ block, participants heard some examples of nonwords containing these phonemes to make it clear, particularly to the Dutch participants, which sounds were intended. The other phonemes were not expected to cause uncertainty, and were not illustrated with examples. Each block started with six practice trials. The response buttons were labeled "A" and "E," "Z" and "S," "V" and "F," "B" and "P," or "D" and "T," respectively. The experiment was controlled with NESU (Nijmegen Experiment Set-Up) experimental software. Stimuli were presented binaurally over Sennheiser closed headphones at a comfortable listening level, one at a time. Participants responded by pressing one of two response buttons. No time limit was imposed for the responses. After each button press, presentation of the next item started.

TABLE IV. Experiment 1, acoustic measures of stimuli with initial and final stops: Proportion of initial stops with prevoicing, mean vowel duration (ms), mean *F*1 offset frequency (Hz), mean closure duration (ms), proportion of final stops with voicing during closure, mean closure voicing duration (as a percentage of total closure duration), and mean burst duration (ms).

	Initial				Final			
	/b/	/p/	/d/	/t/	/b/	/p/	/d/	/t/
Prevoicing	2/3	0	3/3	0	⋯	⋯	⋯	⋯
Vowel duration	⋯	⋯	⋯	⋯	185.7	112.7	252.8	133.3
*F*1 offset frequency	⋯	⋯	⋯	⋯	246	261	272	310
Closure duration	⋯	⋯	⋯	⋯	82.1	117.8	59.1	113.6
Closure voicing	⋯	⋯	⋯	⋯	3/3	0	3/3	3/3
Closure voicing duration (%)	⋯	⋯	⋯	⋯	94.2	0	100	27.5
Burst duration	10.8	22.6	16.3	18.7	67.5	72.0	72.0	122.9

TABLE V. Experiment 1 results: Mean percentage of correct responses as a function of participants' native language and condition.

	Dutch				English			
	Medial	Initial	Final released	Final dereleased	Medial	Initial	Final released	Final dereleased
/æ/	96	⋯	⋯	⋯	97	⋯	⋯	⋯
/ɛ/	94	⋯	⋯	⋯	100	⋯	⋯	⋯
/z/	⋯	96	96	⋯	⋯	97	93	⋯
/s/	⋯	96	98	⋯	⋯	98	95	⋯
/v/	⋯	95	94	⋯	⋯	98	97	⋯
/f/	⋯	92	100	⋯	⋯	98	99	⋯
/b/	⋯	86	96	92	⋯	99	98	96
/p/	⋯	98	98	93	⋯	99	98	96
/d/	⋯	99	94	95	⋯	97	96	97
/t/	⋯	99	98	94	⋯	98	95	93

B. Results and discussion

One response with a reaction time (RT) longer than 10 000 ms due to a technical error was removed. One Dutch subject gave only "v" responses for the /v/–/f/ contrast in final position. All responses of this subject on both /v/–/f/ contrasts were removed from the analysis. Mean percentages of correct responses are presented in Table V. The sensitivity measure d' was calculated for each subject, for each contrast, and each condition separately, with a correction for near-perfect sensitivity (MacMillan and Creelman, 1991). Next, $\log \beta$ was calculated to investigate possible biases (McNicol, 1972). Mean values of d' and $\log \beta$ are presented in Table VI.

For the /æ/–/ɛ/ contrast, an analysis of variance (ANOVA) showed that the d's of the English listeners were significantly larger than those of the Dutch listeners [$F(1,39)=7.59, p<0.01$], indicating a higher sensitivity of the English listeners. However, a t-test showed that the Dutch listeners performed amply above chance ($d'=0$), with d's significantly larger than 3 [$t(19)=5.38, p<0.001$]. There was no effect of native language on bias [$F(1,39)=2.21, p>0.1$].

For the /z/–/s/ contrast, no interaction between condition and native language [$F(1,38)=1.93, p>0.1$], and no main effects of condition [$F(1,38)<1$] and native language [$F(1,38)<1$] were found. No difference in bias between the language groups was found for initial position [$F(1,39)<1$] or for final position [$F(1,39)<1$].

TABLE VI. Experiment 1 results: Mean d' and $\log \beta$ as a function of participants' native language and condition. (Higher values of d' indicate higher sensitivity. Negative values of $\log \beta$ indicate a bias towards the first, and positive values towards the second phoneme of a contrast.)

	Dutch		English	
	d'	$\log \beta$	d'	$\log \beta$
/æ/–/ɛ/	4.21	−0.28	4.96	0.59
Initial /z/–/s/	4.48	−0.25	4.57	0.00
Final /z/–/s/	4.55	0.33	4.17	−0.07
Initial /v/–/f/	4.20	−0.64	4.83	0.00
Final /v/–/f/	4.61	1.23	4.79	0.43
Initial /b/–/p/	3.93	1.76	5.14	0.00
Final /b/–/p/ released	4.68	0.34	4.72	0.02
Final /b/–/p/ dereleased	3.85	0.42	4.31	0.11
Initial /d/–/t/	5.11	−0.02	4.73	0.42
Final /d/–/t/ released	4.45	0.39	4.21	−0.31
Final /d/–/t/ dereleased	3.90	0.09	4.11	−1.00

For the /v/–/f/ contrast, no interaction between condition and native language [$F(1,37)=1.22, p>0.1$], and no main effects of condition [$F(1,37)<1$] and native language [$F(1,37)=2.40, p>0.1$] were found. No difference in bias between the language groups was found for initial position [$F(1,39)=1.07, p>0.1$] or for final position [$F(1,38)=2.27, p>0.1$].

For the /b/–/p/ contrast, a significant interaction between condition and native language was found [$F(2,76)=4.45, p<0.05$]. A planned comparison of initial position and final position with release burst yielded a significant interaction between condition and native language [$F(1,38)=8.65, p<0.01$]. Therefore, separate analyses were performed for both conditions and both language groups. In initial position, the English listeners' sensitivity was significantly higher than the Dutch listeners' sensitivity [$F(1,39)=19.75, p<0.001$]. In final position with release burst, there was no effect of native language [$F(1,39)<1$]. Comparing initial position and final position with release burst for the Dutch listeners only, a significantly lower sensitivity was found for initial position [$F(1,19)=5.68, p<0.05$]. For the English listeners, there was no difference between initial position and final position with release burst [$F(1,19)=2.98, p>0.1$].

In a planned comparison of final position with release burst and final position without release burst, no interaction was found between condition and native language [$F(1,38)=1.21, p>0.1$]. The effect of condition was significant [$F(1,38)=10.69, p<0.01$], with d' being larger for final position with release burst than for final position without release burst. There was no significant effect of native language [$F(1,38)<1$].

For initial position, the effect of native language on bias was significant [$F(1,39)=13.94, p<0.001$], with a bias towards "p" responses for the Dutch listeners, and no bias ($\log \beta=0$) for the English listeners. Neither the analysis of $\log \beta$ for final position with release burst [$F(1,39)<1$] nor that without release burst [$F(1,39)<1$] yielded a significant effect.

The Dutch listeners' bias towards "p" responses in ini-

tial position suggested that their low sensitivity for the initial contrast resulted from a high number of errors on /b/ items rather than on /p/ items. As Table V shows, the Dutch listeners' percentage of correct responses for /p/ items was similar to that of the English listeners (98% vs 99%), whereas the Dutch listeners' percentage of correct responses for /b/ items was only 86%, compared to 99% for the English listeners. Acoustical examination of the /b/ items showed that two of the tokens were produced with prevoicing, and the third without prevoicing (Table IV). The tokens with prevoicing received 99% and 96% correct responses from the Dutch listeners, whereas the token without prevoicing received only 63% correct responses from the Dutch listeners. This score is identical to the percentage of correct responses for Dutch initial voiced stops produced without prevoicing found by Van Alphen and Smits (2004). As expected, the English listeners categorized all tokens of initial voiced stops accurately, regardless of the presence or absence of prevoicing.

Finally, for the /d/–/t/ contrast, no interaction was found between condition and native language [$F(2,76)=1.18, p>0.1$]. There was a significant effect of condition [$F(2,76)=10.54, p<0.001$]. A *posthoc* Bonferroni test showed that the sensitivity scores were higher in initial position (note that all initial /d/'s were prevoiced) than in final position with release burst ($p<0.05$) and in final position without release burst ($p<0.001$), and that the two conditions in final position did not differ significantly from one another ($p>0.1$). No effect of native language was found [$F(1,38)<1$]. There was no effect of native language on bias for initial position [$F(1,39)=1.08, p>0.1$], for final position with release burst [$F(1,39)=1.63, p>0.1$], or for final position without release burst [$F(1,39)=3.75, p=0.060$].

For both Dutch and English listeners, the removal of the release burst affected the categorization of the /b/–/p/ contrast but not of the /d/–/t/ contrast. The signal remaining after removal of the release burst may have contained clearer cues for the alveolar stops than for the bilabial stops. For example, the difference in vowel duration and in $F1$ offset frequency of /d/ and /t/ was larger than that of /b/ and /p/ (Table IV).

In general, the results for the four consonant contrasts show a consistent pattern. For the sensitivity measure, no interactions between position and native language were found for the contrasts /v/–/f/, /z/–/s/, and /d/–/t/. The Dutch listeners' categorization of /v/–/f/ and /z/–/s/ was as accurate in final position as in initial position, and as accurate as that of the English listeners. For the /d/–/t/ contrast, both groups performed better on initial position than on final position with release burst. For the /b/–/p/ contrast, there was an interaction between condition and native language. The English listeners outperformed the Dutch listeners on the initial position. The absence of prevoicing of initial stops hindered the Dutch listeners' categorization in English as much as it does in Dutch. As voiced stops are prevoiced less often in English than in Dutch, this may cause Dutch listeners to misperceive the voicing of initial stops in English more frequently than in Dutch. Dutch and English listeners performed equally well on the /b/–/p/ contrast in final position with release burst. Although Dutch does not allow for voiced obstruents in final, prepausal position, the Dutch listeners did not have a bias towards voiceless responses in final position.

The results are in line with the predictions of the PAM. Whereas the Dutch listeners categorized the /æ/–/ɛ/ contrast less accurately than the English listeners did, they categorized the English final voicing contrasts as accurately as the initial contrasts, and as accurately as the English listeners did.

Experiment 1 also tested categorization accuracy for final stops without a release burst. A difference between the Dutch and English listeners' categorization accuracy could have indicated a differential use of the duration of the preceding vowel as a voicing cue. However, the removal of the release bursts of final stops did not influence the Dutch and the English listeners differentially. For the /b/–/p/ contrast, Dutch and English listeners performed better on items with release burst than on the same tokens without release burst. The removal of the release burst affected the performance of the two language groups to the same extent. For the /d/–/t/ contrast there was no difference in sensitivity to items with or without release burst, neither for the Dutch nor for the English listeners. Thus, experiment 1 did not provide any evidence that the Dutch listeners used vowel duration less than the English listeners did.

However, experiment 1 was not a direct test of the use of vowel duration as a cue. Apart from vowel duration, several other cues remained available after removal of the release burst (e.g., $F1$ offset frequency, closure voicing; see Table IV). Thus, the Dutch listeners may have achieved a native-like level of accuracy without using vowel duration as a cue. On the other hand, the absence of a release burst may have stimulated the Dutch listeners to use vowel duration, while they may not do so when more perceptual cues are available. The results from this experiment are not decisive about these possibilities. In fact, any evidence of Dutch listeners using vowel duration as a cue for final voicing would leave open the possibility that their use of the cue was a reaction to the task at hand.

Therefore, the use of vowel duration as a cue to final obstruent voicing was investigated from a different angle in experiment 2. In this experiment, stimulus materials were constructed such that they did not stimulate but rather discouraged the use of vowel duration as a voicing cue. The question was addressed whether Dutch listeners use vowel duration as a voicing cue as persistently as English listeners do. If the Dutch listeners did not use vowel duration in experiment 2, this would not imply that they never do so. However, it could show that Dutch listeners do not use this cue as persistently as English listeners.

It was argued that among the languages that do not have voiced and voiceless obstruents in final position, Dutch prepares its listeners well for the use of vowel duration as a cue to English word-final obstruent voicing. As Dutch has long and short vowels, Dutch listeners are familiar with the use of phonetic vowel duration. They also have native-language experience with the use of vowel duration as a cue to word-medial obstruent voicing (Van den Berg, 1989). Although the role of vowel duration as a cue to voicing seems to be smaller in Dutch than in English, Dutch listeners' familiarity

with the cue in word-medial position may facilitate its use in word-final position in English. Especially advanced learners of English may have learned to use this word-final voicing cue, combining their native and non-native language experience. Therefore, Dutch listeners with a high level of proficiency in English might be expected to process the English obstruent voicing contrast in a native-like manner, with a native-like use of the vowel duration cue. If Dutch listeners do not use vowel duration in a native-like manner, however, this would raise the question whether non-native listeners can ever be expected to process non-native but familiar phoneme contrasts in unfamiliar positions in a native-like manner.

In experiment 2, Dutch and English listeners' categorization of English final voiced and voiceless obstruents was investigated again. For reasons of comparison, categorization of initial voicing contrasts was tested as well. For practical reasons regarding the construction of phoneme continua, only fricatives were tested.

III. EXPERIMENT 2

A. Method

1. Participants

Twenty-eight native speakers of Dutch and 28 native speakers of British English, none of whom had participated in experiment 1, took part in the experiment. The Dutch participants had a high level of proficiency in English as a second language (as in experiment 1), whereas the English participants did not know any Dutch. The Dutch participants were recruited from the Max Planck Institute participant pool, and the English participants were recruited from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex or at the University of Birmingham. None reported any hearing loss. All were volunteers and received a small fee for participation.

2. Materials

The same nonwords for initial and final fricative contrasts were used as in experiment 1. The materials were recorded by the same native speaker of British English who recorded the materials for experiment 1. The speaker read the items one by one, separated by a pause, in a clear citation style. The materials were recorded with a Sennheiser microphone in a soundproof booth onto digital audiotape and downsampled to 16 kHz during transfer to a computer. For each contrast, two target sounds and one or two carriers were extracted from the sound file, using the speech editor XWAVES. The target sounds were used to create voicing continua which were spliced onto the appropriate carriers, as described below.

From the nonword /zi:f/, /i:f/ was extracted, removing the initial /z/, with the cut being made at the first positive zero crossing after the offset of frication noise. From /fu:s/ and /fu:z/, /fu:/ was extracted, truncating the signal at the last positive zero crossing before the onset of frication noise. In a similar way, /u:k/ was extracted from /vu:k/, removing the initial /v/, and /ku:/ was extracted from /ku:f/ and /ku:v/, removing the final /f/ and /v/. These elements served as carriers.

An initial /z/ was extracted from another token of /zi:f/, truncating the signal at the first positive zero crossing after the offset of the frication noise. A final /z/ was extracted from another token of /fu:z/, with the cut being made at the last positive zero crossing before the start of the frication noise. Similarly, an initial and a final /s/, /v/, and /f/ were extracted from other tokens of /si:f/, /fu:s/, /vu:k/, /ku:v/, /fu:k/, and /ku:f/, respectively. For the initial and final /s/ and /f/, a portion in the center of the fricative was removed, such that the duration of each voiceless fricative matched the duration of its voiced counterpart. The initial /s/ and /f/ were shortened by 2 and 8 ms, respectively, and the resulting durations were 115 ms for the /s/ and 148 ms for the /f/. The final /s/ was shortened by 80 to 187 ms, and the final /f/ was shortened by 56 to 127 ms. The final /s/ and /f/ were shortened by 30% and 31%, respectively. The four pairs of phonemes that were thus obtained served as the endpoints of the four continua. For each continuum, nine intermediate steps were generated, following the procedure of Stevenson (1979) and Repp (1981). In this procedure, the amplitudes of two waveforms are added in varying proportions. The proportions have a ratio of 0–1 and 1–0 in the two endpoints, and are equally spaced in the intermediate steps, always adding up to 1.

For each continuum, the two endpoints and the nine intermediate steps were spliced onto the appropriate carriers. Thus, the resulting stimuli ranged from /zi:f/ to /si:f/ and from /fu:z/ to /fu:s/ for the alveolar fricatives, and from /vu:k/ to /fu:k/ and from /ku:v/ to /ku:f/ for the labiodental fricatives. For the two initial continua, there was one carrier each. For the two final continua, there were two carriers each. One was originally pronounced with a voiceless final fricative and contained a phonetically short vowel (of 118 ms for the /z/–/s/ contrast and 98 ms for the /v/–/f/ contrast); the other was originally pronounced with a voiced final fricative and contained a phonetically long vowel (of 233 ms for the /z/–/s/ contrast and 257 ms for the /v/–/f/ contrast).

3. Design

Stimuli were blocked by contrast, position, and carrier. Each block was presented to half of the participants. As there were two carriers for final contrasts and one for initial contrasts, half of the participants only heard the two final contrasts, and the other half heard the two final contrasts and the two initial contrasts. The order of the blocks with initial and final contrasts was counterbalanced where applicable. As explained above, the nontarget consonants in the CVC items were obstruents. As this restriction yields a limited number of items that are nonwords in both languages, /f/ was part of the carrier in the items where /z/ and /s/ were the target sounds. Therefore, the /v/–/f/ contrast was always tested before the /z/–/s/ contrast. Each block consisted of 20 repetitions of the 11 steps of the continuum, semirandomized such that the same step could not occur twice in succession.

Crucially, each participant was presented with only one carrier for each final contrast. For each participant, the dura-

TABLE VII. Experiment 2 results: Mean percentage of "z" or "v" responses to initial fricatives as a function of the place on an 11-step stimulus continuum ranging from /z/ to /s/ or from /v/ to /f/ and participants' native language.

	Dutch		English	
	/z/–/s/	/v/–/f/	/z/–/s/	/v/–/f/
1 (Voiced)	85	93	95	98
2	83	94	91	96
3	83	94	94	97
4	83	93	92	93
5	83	87	91	88
6	79	72	83	76
7	63	53	78	55
8	56	37	65	27
9	41	23	50	12
10	28	15	26	6
11 (Voiceless)	10	9	8	3

tion of the vowel for each final contrast was unvarying. Thus, vowel duration was not informative for the voicing contrast. For all participants, there was a mismatch between vowel duration and other information in the signal for a subset of the stimuli. When voiced fricatives were preceded by a short vowel, or voiceless fricatives by a long vowel, vowel duration and information in the frication noise pointed in opposite directions.

4. Procedure

The procedure was as described for experiment 1. Each block was preceded by a practice part containing two presentations of each of the 11 steps of the continuum in semirandomized order. The response buttons were labeled "Z" and "S," or "V" and "F," respectively.

B. Results and discussion

Eight responses with RTs longer than 10 000 ms due to a technical error were removed from the analysis. The categorization curves of each contrast in each position and for each subject separately were fitted with logistic regression. From the regression models, 50-percent crossover points were retrieved, reflecting the location of the category boundary. From the models, a measure of the steepness of the categorization curve at the 50-percent crossover point was computed, indicating how categorical perception was. In five response sets, the percentage of correct responses at step 1 or 11 did not exceed 50% (one response set representing one contrast in one position for one subject). No logistic regressions were performed on those response sets.

The categorization results for the initial contrasts are presented in Table VII. There were no differences between the Dutch and English listeners in the steepness of the slopes, either for the /z/–/s/ contrast $[F(1,27)=1.55, p>0.1]$ or for the /v/–/f/ contrast $[F(1,26)<1]$.

The categorization results for the final /z/–/s/ contrast are presented in Fig. 1. If vowel duration was used for the categorization of ambiguous fricatives from the middle region of the voicing continuum, this should have resulted in a shift between the curves corresponding to the long and short vowel conditions. However, an ANOVA on the 50-percent crossover points showed no interaction between vowel duration and native language $[F(1,54)=2.53, p>0.1]$, and no main effects of vowel duration $[F(1,54)=1.17, p>0.1]$ or native language $[F(1,54)<1]$.

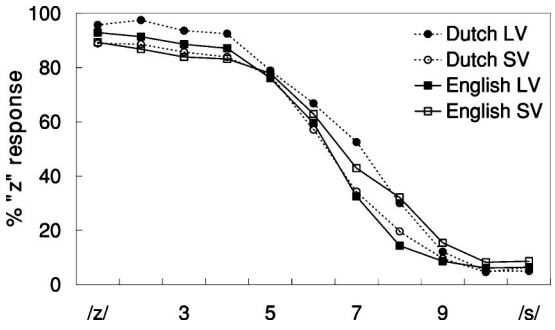


FIG. 1. Experiment 2: Mean percentage of "z" responses to final fricatives as a function of the place on an 11-step stimulus continuum ranging from /z/ to /s/, preceding vowel duration (*LV*: long vowel; *SV*: short vowel), and participants' native language.

The categorization results for the final /v/–/f/ contrast are presented in Fig. 2. The graphs show that vowel duration had a differential effect on the Dutch and the English participants' categorization results. A significant interaction was found between the effects of vowel duration and native language on 50-percent crossover point $[F(1,52)=4.32, p<.05]$.

For the Dutch listeners, the categorization curves were similar in the conditions with the short and with the long preceding vowel. Although the curve for items with a long vowel was located slightly further towards the voiceless side of the continuum than the curve for the short vowel, the 50-percent crossover points were not statistically different $[F(1,27)<1]$.

For the English listeners, the categorization curve for the items with a long vowel was strongly shifted relative to the curve for items with a short vowel. The 50-percent crossover points were significantly different in the two conditions $[F(1,24)=18.63, p<0.001]$, with a larger 50-percent crossover point for the condition with longer vowel duration, showing a preference for "v" responses which persisted further towards the voiceless side of the continuum.

Moreover, the English listeners categorized even the /f/ endpoint as "v" 31% of the time when it was preceded by a long vowel. The difference between the curves for the short

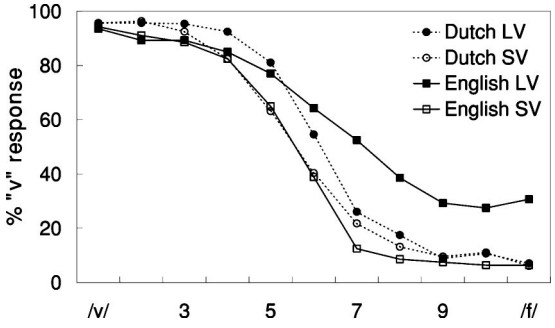


FIG. 2. Experiment 2: Mean percentage of "v" responses to final fricatives as a function of the place on an 11-step stimulus continuum ranging from /v/ to /f/, preceding vowel duration (*LV*: long vowel; *SV*: short vowel), and participants' native language.

and long vowel conditions was located on the voiceless side of the continuum. ANOVAs on arcsine-transformed proportions showed that from steps 6 to 11 the proportion of "v" responses was significantly higher for items with a long vowel than for items with a short vowel ($p<0.01$).

There was no effect of native language on steepness of the slope in the short vowel condition [$F(1,27)<1$]. In the long vowel condition, the Dutch listeners' categorization curve was steeper than the English listeners' curve [$F(1,24)=4.76, p<0.05$], indicating that the Dutch listeners' categorization was more categorical than the native English listeners' categorization.

Vowel duration thus affected the categorization of the /v/–/f/ contrast and the /z/–/s/ contrast differentially. For the final /v/–/f/ contrast, a change in vowel duration led to a shift in the categorization curve for the English but not for the Dutch listeners. For the final /z/–/s/ contrast, there was no shift for either language group. A significant three-way interaction [$F(1,48)=10.12, p<0.001$] among the effects of vowel duration, native language, and place of articulation on 50-percent crossover point confirmed that the contrasts differed in this respect.

The finding that there was no effect of vowel duration for the final /z/–/s/ contrasts is not surprising in itself, as vowel duration was not informative in this experiment. As vowel duration was kept constant for each participant throughout the whole block, it did not have any cue value for the voicing contrast. Nevertheless, the English listeners but not the Dutch listeners showed an effect of vowel duration in their categorization of the final /v/–/f/ contrast. As the English listeners based their categorization decisions on the uninformative vowel duration, their categorization for the /v/–/f/ contrast preceded by a long vowel was less categorical than the Dutch listeners' categorization.

The different results for the alveolar and labiodental fricatives may be caused by their acoustic characteristics. Word-initial alveolar fricatives have a higher amplitude than labiodental fricatives (Jongman, Wayland, and Wong, 2000). A similar difference may exist in final position. Indeed, the spectral power of the final /z/ and /s/ was higher than that of the final /v/ and /f/ in experiment 1 (Table III) and in experiment 2 (where the mean logarithms of the spectral power of the frication noise above 500 Hz in Pa^2 were -3.1 and -4.3, respectively). Further, in final position, alveolar fricatives have a longer noise duration than labiodental fricatives (Crystal and House, 1988). This was also the case in experiment 1 (Table III) and in the original final fricatives in experiment 2. Therefore, the information in the alveolar frication signal may generally be more easily perceptible for the listener than the information in a labiodental. A less informative frication signal may stimulate listeners to exploit other sources of information. This may explain why the English participants took vowel duration into consideration in their decisions for the final /v/–/f/ contrast, but not for the /z/–/s/ contrast.

English listeners categorized the same tokens significantly more often as "v" when they were preceded by a phonetically long vowel than when they were preceded by a short vowel. Even the endpoint /f/ was categorized as "v" 31% of the time in the long vowel condition. Apparently, the long duration of the vowel pointed towards a voiced fricative so strongly that it overruled the other information in the signal in many cases. Note that the reverse did not happen on the other side of the continuum: tokens at the voiced end of the continuum received a high percentage of voiced responses, even when preceded by a short vowel. Several factors may have contributed to this asymmetry. In the first place, the final /f/ was shortened to match the duration of the final /v/. As frication duration is a cue to voicing (e.g., Watson, 1983), the shortening made the final /f/ more /v/-like. Note that this shortening did not lead to a high percentage of "v" responses for the Dutch listeners, or for the English listeners in the short vowel condition. Neither was there a high percentage of "z" responses to the endpoint /s/, which was shortened to a similar extent. Thus, the shortening cannot explain the high percentage of "v" responses for the English listeners in the long vowel condition, but it may have made the /f/ endpoint more acceptable as a "v" than vice versa. Second, listeners have experience with vowel shortening in fast speech. In fast speech, vowels are reduced relatively more than consonants (Gay, 1978), so that not only the vowel duration itself, but also the ratio of vowel and fricative duration changes. Indeed, vowel lengthening before voiced obstruents decreases at faster speaking rates (Smith, 2002). As a result of this experience with absolute and relative vowel shortening, listeners may find it easier to ignore short vowel duration as a cue to voicing when it mismatches with other cues than to ignore long vowel duration. In the third place, phonologically voiced obstruents are phonetically often unvoiced (Stevens *et al.*, 1992). Therefore, listeners may show asymmetric weighting of the presence or absence of phonetic voicing. The presence of phonetic voicing may signal a voiced obstruent relatively strongly, while its absence may not point as strongly towards a voiceless interpretation. This may have contributed to the finding that tokens on the voiced side of the continuum were predominantly perceived as voiced, whereas in the long vowel condition, tokens on the voiceless side of the continuum were less often perceived as voiceless.

IV. GENERAL DISCUSSION

The results of the two experiments presented in this paper show that a native-like level of accuracy may be reached for the categorization of non-native phonemes, even though the phonemes are not necessarily processed in a native-like manner.

Experiment 1 investigated the accuracy with which Dutch listeners categorized English contrasts with different degrees of correspondence in Dutch phonology. The /æ/–/ɛ/ contrast, which the PAM predicts to be among the most difficult English contrasts for Dutch listeners, was indeed found to be the most difficult contrast in this experiment. Although the Dutch listeners performed amply above chance, the English listeners showed a significantly higher sensitivity than the Dutch listeners did. The obstruent voicing contrasts are matched by similar contrasts in Dutch, and the PAM predicts them to be easy to distinguish. Although Dutch voicing contrasts do not occur in final position, the Dutch listeners cat-

egorized the English final voicing contrasts as accurately as (or even more accurately than) the initial contrasts, and as accurately as the English listeners did. Dutch listeners were not biased towards voiceless responses in final position.

The PAM does not make any predictions about the perception of familiar contrasts in unfamiliar positions. The present study suggests that an unfamiliar position does not necessarily complicate the perception of familiar but non-native contrasts. For example, in experiment 1 the Dutch listeners categorized the /b/–/p/ contrast more accurately in the unfamiliar final position than in the familiar initial position. In order to make predictions about the perception of familiar contrasts in familiar and unfamiliar positions, it is important to take into account that speech sounds have different acoustic characteristics in different positions. The extent to which these characteristics overlap with those of the native speech sounds seems an important predictor of the ease with which non-native listeners can distinguish between the sounds.

In experiment 1, no evidence was found that the Dutch listeners used vowel duration as a cue to final voicing less than the English listeners did. The removal of the release burst from final stops did not affect the Dutch listeners' categorization more than it affected the English listeners' categorization. However, the Dutch listeners may have achieved a native-like level of accuracy without using vowel duration as a cue. Flege (1989) found a native-like level of accuracy for Chinese listeners' categorization of unedited tokens of English final /d/–/t/. Nonetheless, the Chinese listeners were found to rely on cues in the release burst more than the English listeners did. They had achieved a native-like accuracy through a non-native-like manner of processing. When the release burst was removed, they were no longer able to maintain a native-like level of accuracy. The Dutch listeners may have been better able to adapt to the removal of the release burst. Their knowledge about English voicing cues may have been sufficient to use those cues that remained available when the burst had been removed (e.g., $F1$ offset frequency, closure voicing). They may have achieved a native-like level of accuracy for the categorization of stops without release burst without using vowel duration. On the other hand, it is also possible that the absence of a release burst stimulated the Dutch listeners to use vowel duration, while they may not do so when more perceptual cues are available. Thus, the results from experiment 1 left the possibility open that non-native listeners do not need to process non-native phonemes in a native-like manner in order to achieve a native-like level of accuracy.

Therefore, experiment 2 tested the use of the duration of the preceding vowel as a cue to final fricative voicing with items which were constructed such that they did not stimulate the use of vowel duration as a voicing cue. In experiment 2, categorization of initial fricative voicing contrasts was tested as well. In line with the results from experiment 1, no differences were found between Dutch and English listeners' categorization of initial fricative voicing contrasts.

In order not to stimulate the Dutch listeners to use vowel duration more than they would normally do, but rather discourage its use, the vowel duration cue was kept uninformative. Vowel duration even mismatched with other information in the signal for some of the tokens. Nevertheless, the English listeners tried to use vowel duration for the categorization of /v/ and /f/. Especially, the categorization of tokens with long vowels was often consistent with vowel duration. Apparently, for the English listeners, vowel duration was such an important cue for final voicing that it often overruled other information in the signal. The Dutch listeners, on the other hand, did not use vowel duration at all. Even in the middle range of the continuum, for tokens with an ambiguous identity, there was no effect of vowel duration on the Dutch listeners' categorization of final /z/–/s/ or /v/–/f/. Thus, in experiment 2, Dutch listeners did not categorize final voiced and voiceless obstruents in a native-like manner. The Dutch listeners were able to ignore vowel duration when it was uninformative and misleading. They differed in this respect from the English listeners, who ignored vowel duration for the categorization of the final /z/–/s/ contrast, but relied heavily on it for the categorization of the final /v/–/f/ contrast. As a result, for the /v/–/f/ contrast preceded by a long vowel, Dutch listeners' categorization curve was steeper than that of the native English listeners. As vowel duration was not informative and sometimes mismatched with other voicing cues, English listeners' use of vowel duration for the /v/–/f/ contrast resulted in less categorical perception.

The Dutch listeners, who did not use vowel duration as a cue in experiment 2, may do so in other circumstances where the cue is informative. Indeed, they may have used vowel duration as a cue in experiment 1. However, the results from experiment 2 showed that the Dutch listeners did not use vowel duration as persistently as the English listeners did. This may be a result from their native-language experience, where vowel duration is a less important cue to (word-medial) obstruent voicing than in English. From their native-language experience, the Dutch listeners may have inferred that vowel duration is only a minor cue to English final obstruent voicing as well. Another explanation could be that Dutch listeners are regularly exposed to English spoken by native speakers of Dutch. Elsendoorn (1985) has shown that the difference in vowel duration before voiced and voiceless final obstruents in the English spoken by Dutch learners is smaller than that in the English of native speakers. From exposure to English spoken by Dutch learners, Dutch listeners may have learned to ignore vowel duration as a voicing cue when it is uninformative.

It was argued that Dutch prepares its listeners well for the distinction of English word-final obstruent voicing. Dutch has obstruent voicing contrasts which are perceptually similar to English contrasts, which makes the English contrasts easy to distinguish according to the PAM (Best *et al.*, 1988). Dutch allows for voiceless obstruents in word-final position, which has been found to facilitate perception of the distinction (Flege and Wang, 1989). Dutch provides experience with phonetic vowel duration for the distinction of phonemically long and short vowels, which has been found to facilitate the use of this cue for the final consonant voicing contrast (Crowther and Mann, 1992), and Dutch provides experience with the use of vowel duration as a cue to word-medial obstruent voicing (Van den Berg, 1989). Indeed,

Dutch listeners were found to categorize English final obstruent voicing with a native-like level of accuracy. Nevertheless, they were found to use vowel duration as a cue less persistently than the English listeners did.

As even native listeners of Dutch, who had the necessary experience for a native-like use of vowel duration, and who had a high level of proficiency in English, did not use perceptual cues in a native-like manner, this raises the question whether non-native listeners can ever be expected to process non-native but familiar phoneme contrasts in unfamiliar positions in a native-like manner.

Although the Dutch listeners in this study had a high level of proficiency in English, they had not reached a level of ultimate attainment. Their English perception skills were still open to improvement. However, such improvement may not involve the use of perceptual cues for contrasts which they could already accurately distinguish. As the results from the experiments in this paper show, a native-like level of accuracy can be achieved, even when the perceptual cues are not processed in a native-like manner. Possibly, Dutch listeners can learn to use vowel duration as a cue to English word-final obstruent voicing in a fully native-like manner with a native-like persistence, for example through laboratory-based training, but there may be no need to learn this for normal language use. Presumably, the second language learner's goal is not to process language in a native-like manner, but rather to be able to understand (and produce) the language well enough to meet the learner's communicative needs. If a native-like use of perceptual cues is not necessary for accurate perception, it is possible that listeners may never learn to perceive non-native phonemes in a native-like manner. The benefits of native-like processing may be too small, or even nonexistent.

It should be relatively easy for Dutch listeners to learn to use vowel duration in an English native-like manner, and nonetheless, in experiment 2 the Dutch listeners did not use vowel duration as a cue when the English listeners did. As even these Dutch listeners did not use vowel duration in a native-like way, it seems likely that native listeners of languages for whom the cue is harder to learn would in many cases not use the cue in a native-like manner either. The reasoning can be extended to other perceptual cues as well. The vowel duration cue provides a great amount of information about the English voicing contrast, and it is a very important cue for English listeners. It therefore seems a good candidate for native-like use by non-native listeners. As even this cue was not used in a native-like way, it seems likely that other, less informative cues would in many cases not be used in a native-like manner either. Thus, for the distinction of non-native but familiar contrasts in unfamiliar positions, a native-like manner of phonetic processing may hardly ever be attained.

ACKNOWLEDGMENTS

This work was supported by a grant from the Max Planck Society for the Advancement of Science and by a Marie Curie fellowship. I thank Antje Meyer of the Behavioral Brain Science Center, University of Birmingham, and Alan Garnham of the Laboratory of Experimental Psychology, University of Sussex, for giving me the opportunity to work in their labs. Many thanks to Anne Cutler, James McQueen, and three anonymous reviewers for their helpful comments on earlier versions of this manuscript.

Baayen, H., Piepenbrock, R., and Gulikers, L. (**1995**). *The* CELEX *Lexical Database (CD-Rom)* (Linguistic Data Consortium, University of Pennsylvania, Philadephia, PA).

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–224.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). ''Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants,'' J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Best, C. T., and Strange, W. (**1992**). ''Effects of phonological and phonetic factors on cross-language perception of approximants,'' J. Phonetics **20**, 305–330.

Booij, G. (**1995**). *The Phonology of Dutch* (Oxford University Press, Oxford).

Byrd, D. (**1993**). ''54 000 American stops,'' UCLA Working Papers **83**, 97–115.

Crowther, C. S., and Mann, V. (**1992**). ''Native language factors affecting use of vocalic cues to final consonant voicing in English,'' J. Acoust. Soc. Am. **92**, 711–722.

Crystal, T. H., and House, A. S. (**1988**). ''A note on the durations of fricatives in American English,'' J. Acoust. Soc. Am. **84**, 1932–1935.

Elsendoorn, B. A. G. (**1985**). ''Production and perception of Dutch foreign vowel duration in English monosyllabic words,'' Lang Speech **28**, 132–254.

Flege, J. E. (**1989**). ''Chinese subjects' perception of the word-final English /t/–/d/ contrast: Performance before and after training,'' J. Acoust. Soc. Am. **86**, 1684–1697.

Flege, J. E., and Hillenbrand, J. (**1987**). ''A differential effect of release bursts on stop voicing judgments of native French and English listeners,'' J. Phonetics **15**, 203–208.

Flege, J. E., and Wang, C. (**1989**). ''Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/–/d/ contrast,'' J. Phonetics **17**, 299–315.

Gay, T. (**1978**). ''Effect of speaking rate on vowel formant movements,'' J. Acoust. Soc. Am. **63**, 223–230.

Jongman, A., Sereno, J. A., Raaijmakers, M., and Lahiri, A. (**1992**). ''The phonological representation of [voice] in speech perception,'' Lang. Speech **35**, 137–152.

Jongman, A., Wayland, R., and Wong, S. (**2000**). ''Acoustic characteristics of English fricatives,'' J. Acoust. Soc. Am. **108**, 1252–1263.

Kuijpers, C. T. L. (**1996**). ''Perception of the voicing contrast by Dutch children and adults,'' J. Phonetics **24**, 367–382.

Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.

MacMillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

McNicol, D. (**1972**). *A Primer of Signal Detection Theory* (Australasian, Sydney, Australia).

Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.

Raphael, L. J. (**1972**). ''Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English,'' J. Acoust. Soc. Am. **51**, 1296–1303.

Repp, B. H. (**1981**). ''Perceptual equivalence of two kinds of ambiguous speech stimuli,'' Bull. Psychon. Soc. **18**, 12–14.

Schouten, M. E. H. (**1975**). ''Native-Language Interference in the Perception of Second-Language Vowels: An Investigation of Certain Aspects of the Acquisition of a Second Language,'' Doctoral dissertation, Utrecht University.

Slis, I. H., and Cohen, A. (**1969a**). ''On the complex regulating the voiced-voiceless distinction. I,'' Lang. Speech **12**, 80–102.

Slis, I. H., and Cohen, A. (**1969b**). ''On the complex regulating the voiced-voiceless distinction. II,'' Lang. Speech **12**, 137–155.

Slis, I. H., and Van Heugten, M. (**1989**). ''Voiced–voiceless distinction in Dutch fricatives,'' in *Linguistics in The Netherlands 1989*, edited by H.

Bennis and A. v. Kemenade (Foris, Dordrecht, Netherlands), pp. 123–132.
Smith, B. L. (**1978**). ''Effects of place of articulation and vowel environment on 'voiced' stop consonant production,'' Glossa **12**, 163–175.
Smith, B. L. (**2002**). ''Effects of speaking rate on temporal patterns of English,'' Phonetica **59**, 232–244.
Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (**1992**). ''Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters,'' J. Acoust. Soc. Am. **91**, 2979–3000.
Stevenson, D. C. (**1979**). ''Categorical Perception and Selective Adaptation Phenomena in Speech,'' Doctoral dissertation, University of Alberta, 1979.
Trubetzkoy, N. S. (**1977**). *Grundzuege der Phonologie*, 6th ed. (Van den Hoeck and Ruprecht., Goettingen [original work published 1939]).
Van Alphen, P. M., and Smits, R. (**2004**). ''Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing,'' J. Phonetics **32**, 455–491.
Van den Berg, R. J. H. (**1989**). ''Perception of voicing in Dutch two-obstruent sequences: Covariation of voicing cues,'' Speech Commun. **8**, 17–25.
Watson, I. (**1983**). ''Cues to the voicing contrast: A survey,'' Cambridge Papers in Phonetics and Experimental Linguistics 2.

Fine structure spectrography and its application in speech

Hilmi R. Dajani,[a] Willy Wong, and Hans Kunov
Institute of Biomaterials and Biomedical Engineering and the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Rosebrugh Building, 4 Taddle Creek Road, Toronto, Ontario M5S 3G9, Canada

(Received 11 June 2002; accepted 18 January 2005)

A filterbank-based algorithm for time-varying spectral analysis is proposed. The algorithm, which is an enhanced realization of the conventional spectrogram, consists of hundreds or thousands of highly overlapping wideband filter/detector stages, followed by a peak detector that probes the filter/detector outputs at very short time intervals. Analysis with synthetic modulated signals illustrates how the proposed method demodulates these signals. The resulting spectrogram-like display, referred to as a ''fine structure spectrogram,'' shows the fine structure of the modulations in substantially higher detail than is possible with conventional spectrograms. Error evaluation is performed as a function of various parameters of a single- and two-component synthetic modulated signal, and of parameters of the analysis system. In speech, the fine structure spectrogram can detect small frequency and amplitude modulations in the formants. It also appears to identify additional significant time-frequency components in speech that are not detected by other methods, making it potentially useful in speech processing applications.
[DOI: 10.1121/1.1896365]

PACS numbers: 43.72.Ar [DOS] Pages: 3902–3918

I. INTRODUCTION

The analysis of signals whose frequency content varies with time has traditionally been performed with the spectrogram (Koenig *et al.*, 1946; Rosen and Howell, 1991). Called the ''workhorse'' of time-frequency analysis (Loughlin, 1999), the spectrogram today can be generated digitally using the short time Fourier transform (STFT). In the case of the speech signal, the spectrogram has been widely used to investigate the features of speech in the time-frequency plane that are perceptually important (Potter *et al.*, 1947; Pitton *et al.*, 1996). In one early and notable application, the spectrogram was used to determine the average frequencies of the first three formants of the vowels of American English for men, women, and children (Peterson and Barney, 1952). A drawback of the spectrogram, however, is the unavoidable tradeoff between time and frequency resolution, also known as the uncertainty principle of signal analysis (Gabor, 1946; Cohen, 1995). When the spectrogram of speech is implemented using the STFT, a short analysis window may result in a poor spectral representation, while a longer window may misrepresent the consonant-vowel transitions (Pitton *et al.*, 1996). Representations based on the wavelet transform, which is thought to provide a closer approximation to auditory frequency analysis than the STFT, are also limited by the uncertainty principle (Quatieri, 2002).

When evaluating the spectrogram, an important assumption is that the signal is stationary over the analysis window—a condition known as quasi-stationarity. Although speech is in reality a nonstationary signal, the assumption of quasi-stationarity in speech processing has been of primary importance because it is closely linked to the source-filter model of speech production (Rabiner and Schafer, 1978; Pitton *et al.*, 1996; Rao and Kumaresan, 2000). This model, which has dominated the field of speech for a long time, assumes that the slowly varying parameters of the filter formed by the vocal tract are the primary carriers of information—at least for nontonal languages such as English. The source signal and its higher frequency harmonics are usually of little interest other than for tracking the pitch of the speaker. Methods such as linear predictive coding and cepstral analysis, which constitute the state-of-the-art in speech processing (O'Shaughnessy, 2000; Rao and Kumaresan, 2000; Quatieri, 2002), are based on the source-filter model and are used to separate the slowly varying filter spectral envelope from the source signal. And because these methods are based on the assumption of signal quasi-stationarity, they all have poor joint time-frequency resolution.

Recently, there has been increasing recognition of the existence of amplitude and frequency modulations in speech that occur on the order of a pitch cycle, and which are usually smeared by quasi-stationary methods of analysis (O'Shaughnessy, 2000; Quatieri, 2002). These modulations are referred to as ''fine structure,'' and originate from time-varying linear mechanisms, from nonlinear interactions between the source and filter, and from nonlinear air flow in the vocal tract (Maragos *et al.*, 2002; Quatieri, 2002). Quatieri argues that this fine structure probably plays a significant role in human perception, and that therefore there is a need for analysis methods with greater time resolution than quasi-stationary methods. Potamianos and Maragos (1999) noted that the fine structure of speech may be the reason why natural speech is more intelligible than synthetic speech under noisy conditions. Rao and Kumaresan (2000) contend that extracting the modulations of speech may shed new light on speech signals, and on speaker-specific information in par-

[a]Author to whom correspondence should be addressed. Electronic mail: h.dajani@utoronto.ca

 0001-4966/2005/117(6)/3902/17/$22.50

ticular. A better understanding of the modulation patterns in speech, therefore, may find application in speech and speaker recognition, and in better quality speech synthesizers and coders.

As part of the effort to obtain better time-frequency resolution, several representations have been developed that do not rely on the assumption of stationarity over short time intervals (Cohen, 1995; Pitton *et al.*, 1996). The most popular of these so-called nonstationary time-frequency representations are different types of bilinear—or Cohen's class—distributions. Bilinear distributions, however, often suffer from artifacts and from regions of negative spectral energy that have no obvious physical meaning (Pitton *et al.*, 1996). For speech analysis, there has been hope that these nonstationary distributions would replace the STFT (Pitton *et al.*, 1996); however, in the case of a complex signal such as speech, they have often been difficult to interpret (Kumaresan and Rao, 1999). So far, all the time-frequency representations that have been discussed characterize the power density of a signal as a function of time and frequency. Therefore, they can be called ''power'' representations. In the case of the speech signal, a very different approach to time-frequency analysis has been recently developed. This approach aims to decompose speech into a set of modulated components (Maragos *et al.*, 1993; Kumaresan and Rao, 1999; Nelson, 2001). The key difference between ''power'' and ''modulation decomposition'' representations is related to how they characterize local time-frequency properties. A power representation aims to represent the local spectral content of a signal, whereas a modulation decomposition representation tracks the instantaneous frequency (and amplitude) of a modulated component. In the case of speech, modulation decomposition methods, which are typically model-based, have been successful in detecting fine structure in the amplitude and frequency modulations of speech. In this paper, we will show how a filterbank-based method can also detect the fine structure in the modulation patterns of speech. Moreover, because it is not model-based, it also appears to detect significant time-frequency modulations not seen with other demodulation methods.

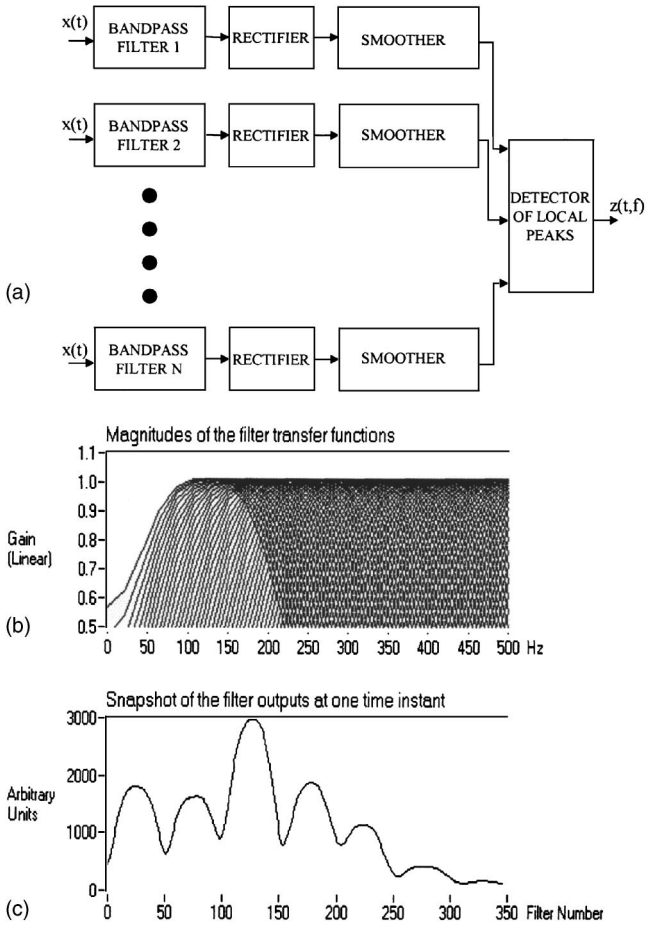


FIG. 1. (a) Basic structure used to generate the fine structure spectrogram (FSS). Hundreds or thousands of stages are used. (b) The magnitudes of the filter transfer functions in one implementation where the center frequencies were separated by 5 Hz. (c) A plot of the rectified and smoothed filter outputs, at a given instant in time. The analyzed signal was the syllable /ba/.

II. FINE STRUCTURE SPECTROGRAM

A. Basic structure

It is possible to formulate a filterbank-based approach that is motivated by one aspect of peripheral auditory processing (Fig. 1). Hundreds or thousands of overlapping filter/detectors (F/D's) are used, inspired by the thousands of tuning characteristics of the afferent auditory fibers (Harrison, 1988). Each F/D consists of a bandpass filter, followed by a rectifier and smoother, with a detector of local peaks in the outputs of all the F/D's as a final stage. The bandpass filter is unimodal and symmetric, and the peak detector fits a quadratic polynomial to sequential groups of the output of three adjacent F/D's and searches for a peak. If a peak is found, it skips another point. When wideband filters are used, the peaks are typically broad when plotted against the filter center frequencies [Fig. 1(c)] and so readily discernable. In the case of speech, these peaks track the prominent resonances of the signal. High time-frequency responsiveness is achieved by having fairly wideband individual filters that respond rapidly to changes in the signal, and filters that are separated by a very fine amount. Responsive tracking of the frequency components also requires that the F/D's be probed at very small time intervals. Peak picking on the outputs of a very large number of filter/detector stages, together with the use of time-responsive wideband filters, allows the algorithm to track the instantaneous frequency (and amplitude) of a modulated component as it evolves in time. A plot of the peaks of the F/D's against the center frequency of the F/D's and against each time instant gives a spectrogram-like display, which is referred to as a ''fine structure spectrogram'' or FSS.

The structure of the FSS is not fundamentally new and can be seen as an enhanced realization of the conventional spectrogram. The original analog spectrograph was composed of several F/D stages (Rosen and Howell, 1991), and the digital implementation of the spectrogram using the STFT has a filterbank interpretation (Rabiner and Schafer, 1978; Oppenheim *et al.*, 1999). Moreover, peak picking on the output of a filterbank (or on the STFT in the case of a digital implementation of the spectrogram) has been used before. Flanagan (1956) proposed the use of peak picking on

the outputs of a bank of a few analog filters as a method to track speech formants. Smits (1994) used peak picking on the wideband STFT to estimate formant frequencies. Peak picking on the STFT has also been used to track the instantaneous frequency of signals, but in the case of rapidly varying frequency modulated (FM) signals, it has been found to give poor results (Boashash, 1992b). What is different about the proposed method is the use of a very large number of wideband overlapping filter/detectors, and the emphasis on the very frequent probing of the outputs of the F/D's. If implemented using the STFT (Sec. II D), very fine sampling in time and frequency (along with a very short window) is used. It is proposed that satisfying these requirements, combined with peak detection, allows the FSS to detect the time-frequency modulation fine structure in a signal such as speech.

Shamma and colleagues used differential filters to remove the redundancy in the outputs of highly overlapping ''cochlear'' filters and to emphasize the time-varying peaks in the spectrum (Yang *et al.*, 1992). In the case of Shamma's approach, the spectral resolution of the resulting representation is determined by the bandwidth of the differential filters and not of the cochlear filters. In order to achieve fine spectral resolution, Shamma uses highly asymmetrical cochlear filters that are relatively broadband (for good dynamic response) but have a sharp edge to allow the differential filters to enhance the spectral peaks. The phase responses of the ''cochlear'' filters may be designed to create discontinuities along the frequency axis, which has been suggested to be the mechanism by which lateral inhibitory networks enhance spectral peaks in the pattern of activity across the auditory fiber array (Shamma, 1985). In the case of FSS, in contrast, the filters used are symmetric and the time relations between the different frequency components of the signal should be preserved at the output of the filterbank by using zero phase filters or filters having linear-phase response with the same slope.

B. Filter/detector (F/D) theory

Anderson (1984) studied in detail the signal processing behavior of various types of F/D's. He concluded that an F/D consisting of a bandpass filter followed by a rectifier and a smoothing filter is equivalent under many conditions to an F/D consisting of a bandpass filter followed by a square-law device and a smoothing filter, with a square-root device attached at the end to rescale the level of the output. This is a fortunate result because the signal processing function of a rectifier is difficult to analyze, as it produces a complex broadband output to simple inputs. In contrast, the signal processing performed by an F/D that incorporates a square-law device is relatively easy to analyze—at least with some inputs.

C. Performance with synthetic modulated signals

In the next four sections (II C 1–II C 4), input signals are taken through the processing stages of the proposed algorithm. The effect of each stage is determined, and then the overall output of the algorithm—the fine structure spectrogram—is established. The terminal square-root device is ignored in all cases. In the next two sections (II C 5 and II C 6), the FSS of two other types of modulated input signals is determined without calculating the intermediate stages. The goal of this analysis is to see if the FSS correctly detects the time-frequency fine structure in a set of synthetic signals with various types of amplitude and frequency modulation that are of relevance to speech.

1. Case of a pure sinusoid

When the input signal is a pure sinusoid $A_0 \sin \omega_s t$,

Output of square-law device:

$$A_1^2(1-\cos 2\omega_s t)/2, \tag{1}$$

where ω_s is the frequency of the sinusoid and A_1 is its amplitude at the output of the bandpass filter.

Output of smoothing filter:

$$A_1^2/2. \tag{2}$$

if the cutoff frequency of the smoothing filter is $<2\omega_s$, and the filter has a sharp response characteristic that effectively suppresses the component at $2\omega_s$. If the bandpass filters are unimodal and symmetric, the F/D whose center frequency coincides with ω_s will have the highest output for all time. Therefore peak picking on the outputs of multiple F/D's and plotting the FSS will result in a horizontal line at ω_s, as is desired.

2. Case of anplitude modulated (AM) signal

A sinusoidal carrier-added AM signal can be represented as (Carden *et al.*, 2002)

$$0.5mA_0 \sin(\omega_c-\omega_m)t+0.5mA_0 \sin(\omega_c+\omega_m)t+A_0 \sin \omega_c t,$$

where ω_c and ω_m are the carrier and modulation frequencies in rad/s, A_0 is the amplitude of the carrier, and m is the modulation index.

Output of bandpass filter:

$$A_1 \sin(\omega_c-\omega_m)t+A_2 \sin(\omega_c+\omega_m)\mathrm{t}+A_3 \sin \omega_c t, \tag{3}$$

where A_1, A_2, and A_3 are the amplitudes of the three components after passing through the (as yet unspecified) bandpass filter.

Output of square law device (after manipulation with simple trigonometric identities):

$$(A_1^2+A_2^2+A_3^2)/2+A_3(A_1+A_2)\cos \omega_m \mathrm{t}+A_1A_2 \cos 2\omega_m \mathrm{t} + \cdots \text{ (components at } 2\omega_c-2\omega_m, 2\omega_c-\omega_m, 2\omega_c, 2\omega_c+\omega_m, 2\omega_c+2\omega_m) \tag{4}$$

Output of smoothing filter: If the cutoff frequency of the smoothing filter is chosen to be $<2\omega_m$ and $\omega_c \gg \omega_m$, the output of a single filter/detector stage is

$$(A_1^2+A_2^2+A_3^2)/2+A_3(A_1+A_2)\cos \omega_m t. \tag{5}$$

Detection of local peaks in the outputs of F/D stages: To find the local peaks in the output of the F/D stages at a given time instant, Eq. (5) is differentiated with respect to the center frequency of the F/D stage. For the purposes of the analy-

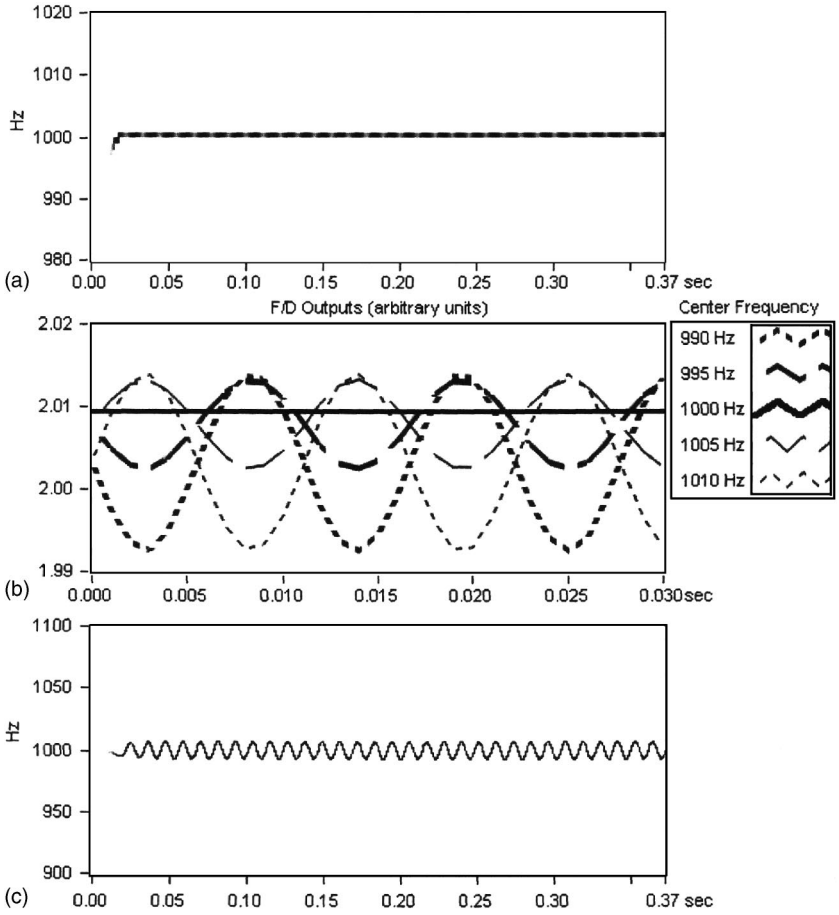


FIG. 2. (a) Fine structure spectrogram of an AM signal with a carrier frequency of 1000 Hz, a modulation frequency of 90 Hz, and a modulation index of 0.25, obtained with a direct computer implementation (bandpass filter bandwidth 200 Hz, low-pass filter cutoff frequency 100 Hz, filter separation 1 Hz, probing interval 0.045 ms). (b) Calculated outputs of F/D stages centered at different frequencies, in arbitrary units, with a narrowband frequency modulated (NBFM) input signal. The output of the F/D whose center frequency coincides with the carrier frequency is the line with no gaps. The input signal has a carrier frequency of 1000 Hz, a modulation frequency of 90 Hz, and a peak frequency deviation of 10 Hz. The bandpass filter is Gaussian with a bandwidth of approximately 400 Hz. (c) Fine structure spectrogram of the NBFM signal obtained with a direct computer implementation (bandpass filter bandwidth 200 Hz, low-pass filter cutoff frequency 100 Hz, filter separation 1 Hz, probing interval 0.045 ms).

sis, the bandpass filter is chosen to be Gaussian, because of its convenient unimodal and symmetric shape:

$$\text{Transfer function of bandpass filter} = e^{(-(\omega-\text{CF})^2/2\sigma^2)}, \tag{6}$$

where ω is the frequency variable in rad/s, CF is the center frequency of the filter, and σ (standard deviation) determines the width of the filter.

So, using the Gaussian filter, the derivative of Eq. (5) with respect to CF is

$$2(A_1A_1'+A_2A_2'+A_3A_3')+A_3'(A_1+A_2)\cos\omega_m t + A_3(A_1'+A_2')\cos\omega_m t, \tag{7}$$

where

$$A_1' = A_1(\omega_c-\omega_m-\text{CF})/\sigma^2,$$

$$A_2' = A_2(\omega_c+\omega_m-\text{CF})/\sigma^2,$$

$$A_3' = A_3(\omega_c-\text{CF})/\sigma^2.$$

By inspection, the derivative [Eq. (7)] is zero when CF $=\omega_c$. Therefore, the F/D centered at the carrier frequency of the AM signal will have the maximum and/or minimum output for all time. Whether it is maximum for all time depends on the choice of the parameters of the signal and analysis system. To given an example, this occurs when the carrier and modulation frequencies are 1000 and 90 Hz, respectively, the modulation index is 0.25, and the σ parameter of the filter is chosen so that the bandwidth of the Gaussian bandpass filter is approximately 200 Hz. In that case, the FSS would correctly represent the signal as a single line whose gray scale varies with the modulation cycle [Fig. 2(a)].

3. Case of narrowband frequency modulated (NBFM) signal

A sinusoidal frequency modulated (FM) signal can be written as

$$A_0\cos(\omega_c t + m\sin\omega_m t), \tag{8}$$

where ω_c and ω_m are the carrier and modulation frequencies in rad/s, A_0 is the amplitude of the carrier, and m is the modulation index which specifies the ratio of the peak frequency deviation to the modulation frequency. An FM signal consists of an infinite number of frequency components. However, if $m \ll \pi/2$ (usually if $m<0.2$), then it is termed narrowband and can be approximated by a carrier and two sidebands (Carden *et al.*, 2002):

$$-0.5mA_0\sin(\omega_c-\omega_m)t + 0.5mA_0\sin(\omega_c+\omega_m)t + A_0\sin\omega_c t. \tag{9}$$

The frequency content of the NBFM signal is very similar to that of an AM signal but the first sideband is 180° out of phase with the two other components.

Output of bandpass filter:

$$-A_1 \sin(\omega_c-\omega_m)t+A_2 \sin(\omega_c+\omega_m)t+A_3 \sin \omega_c t, \tag{10}$$

where A_1, A_2, and A_3 are the amplitudes of the three components at the output of the bandpass filter.

Output of square-law device:

$$(A_1^2+A_2^2+A_3^2)/2+A_3(A_2-A_1)\cos \omega_m t-A_1A_2 \cos 2\omega_m t + \cdots \text{ (components at } 2\omega_c-2\omega_m, 2\omega_c-\omega_m, 2\omega_c, 2\omega_c + \omega_m, 2\omega_c+2\omega_m). \tag{11}$$

Output of smoothing filter: If the cutoff frequency of the smoothing filter is $<2\omega_m$ and $\omega_c \gg \omega_m$, the output of a single filter/detector stage is

$$(A_1^2+A_2^2+A_3^2)/2+A_3(A_2-A_1)\cos \omega_m t. \tag{12}$$

To find the F/D's that yield maximal output as a function of time, this expression is differentiated and the center frequency is solved as a function of time. It can be shown that the output of the FSS is simply the derivative of the phase of the original signal as would be expected if the signal is correctly demodulated. Please see the Appendix for a detailed derivation.

We also explored this problem numerically by passing the signal through the FSS. The outputs of the F/D's that are centered on different frequencies are plotted with Eq. (12) as a function of the parameters ω_c, ω_m, and σ of the Gaussian filter. For example, Fig. 2(b) shows the outputs of five F/D's calculated using Eqs. (6), (9), and (12) with a test NBFM signal having a 1000-Hz carrier frequency, a 90-Hz modulation frequency, and a peak frequency deviation of 10 Hz. The individual F/D stages are centered on frequencies ranging from 990 to 1010 Hz in steps of 5 Hz. The Gaussian bandpass filter has a bandwidth of approximately 400 Hz. With an NBFM input, the individual F/D's alternate in having the largest output during the modulation cycle. This is what allows the ''demodulation'' of an input NBFM signal, as shown in the resulting FSS of Fig. 2(c).

4. Case of combined AM/NBFM signal

If an input signal is jointly modulated in amplitude and frequency at the same modulation frequency—as would be expected for the case of a speech component that is modulated at the pitch frequency (Maragos *et al.*, 1993; Potamianos and Maragos, 1999)—then it can be written as

$$A_0(1+m_{\mathrm{AM}} \cos \omega_m t)\cos(\omega_c t+m_{\mathrm{FM}} \sin \omega_m t), \tag{13}$$

where ω_c and ω_m are the carrier and modulation frequencies in rad/s, m_{AM} is the AM modulation index, m_{FM} is the FM modulation index, and A_0 is the amplitude of the carrier. In this case the individual components of the FM signal can be thought of as being individually amplitude modulated. If the frequency modulation is narrowband, then the signal can be approximated by

$$0.5(m_{\mathrm{AM}}-m_{\mathrm{FM}})A_0 \sin(\omega_c-\omega_m)t+0.5(m_{\mathrm{AM}}+m_{\mathrm{FM}}) \times A_0 \sin(\omega_c+\omega_m)t+A_0 \sin \omega_c t - 0.25m_{\mathrm{AM}}m_{\mathrm{FM}}A_0 \sin(\omega_c-2\omega_m)t + 0.25m_{\mathrm{AM}}m_{\mathrm{FM}}A_0 \sin(\omega_c+2\omega_m)t. \tag{14}$$

Output of bandpass filter:

$$A_1 \sin(\omega_c-\omega_m)t+A_2 \sin(\omega_c+\omega_m)t+A_3 \sin \omega_c t + A_4 \sin(\omega_c-2\omega_m)t+A_5 \sin(\omega_c+2\omega_m)t, \tag{15}$$

where A_1, A_2, A_3, A_4, and A_5 are in this case the signed amplitudes of the five input components at the output of the bandpass filter.

Output of square-law device:

$$(A_1^2+A_2^2+A_3^2+A_4^2+A_5^2)/2+(A_1A_3+A_2A_3+A_1A_4 + A_2A_5)\cos \omega_m t+A_3(A_4+A_5)\cos 2\omega_m t + (A_1A_5+A_2A_4)\cos 3\omega_m t+A_4A_5 \cos 4\omega_m t + \cdots \text{ (components at } 2\omega_c-4\omega_m, 2\omega_c-3\omega_m, 2\omega_c - 2\omega_m, 2\omega_c-\omega_m, 2\omega_c, 2\omega_c+\omega_m, 2\omega_c+2\omega_m, 2\omega_c + 3\omega_m, 2\omega_c+4\omega_m). \tag{16}$$

Output of smoothing filter: If the cutoff frequency of the smoothing filter is $<2\omega_m$ and $\omega_c \gg \omega_m$, then the output of an individual F/D stage is given by

$$(A_1^2+A_2^2+A_3^2+A_4^2+A_5^2)/2+(A_1A_3+A_2A_3+A_1A_4 + A_2A_5)\cos \omega_m t. \tag{17}$$

Again the output of the low-pass filter is differentiated and the center frequency solved as a function of time. The results are identical to case 3 (narrowband FM) in that the output of the FSS is simply the time derivative of the phase function of the original signal. Moreover, we can solve for the amplitude of the output to show that the original AM component is also recovered. Please see the Appendix.

When this problem is solved numerically, we obtain Fig. 3. Figure 3 shows the outputs of five F/D's calculated using Eqs. (6), (14), and (17) with a test AM/NBFM signal having a carrier frequency of 1000 Hz, a modulation frequency of 90 Hz, a peak frequency deviation of 10 Hz, an AM modulation index of 0.25, and an FM modulation index of approximately 0.1. The Gaussian bandpass filter has a bandwidth of approximately 400 Hz. Figures 3(b) and (c) are close-ups of the peaks and troughs of the cycle in Fig. 3(a). During the peak of the cycle, the F/D's centered at 1005 and 1010 Hz have the highest outputs. In contrast, during the trough of the cycle, the F/D centered at 990 Hz has the highest output. Between the peak and the trough, the F/D's centered at the other listed frequencies alternately produce the largest outputs (not shown in close-up). Figure 3 illustrates that not only do the individual F/D's alternate in having the largest output during the modulation cycle, but that these outputs also vary in amplitude during this cycle. This makes it possible for the algorithm to simultaneously demodulate both the frequency and amplitude components of the modulation. Figure 4(a) shows the resulting FSS. The

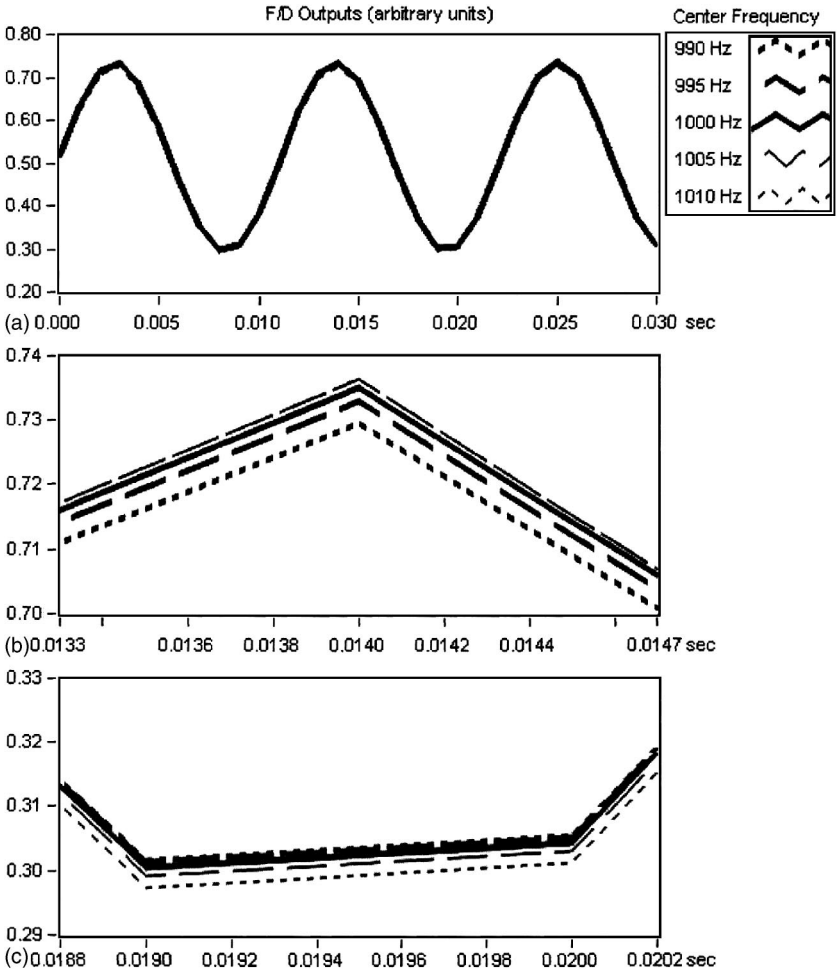


FIG. 3. (a) Calculated outputs of F/D stages centered at different frequencies with a combined amplitude modulated and narrowband frequency modulated (AM/NBFM) input signal. The vertical axis is in arbitrary units (input signal units)^2. The output of the F/D whose center frequency coincides with the carrier frequency is the line with no gaps. The input signal has a carrier frequency of 1000 Hz, a modulation frequency of 90 Hz, a peak frequency deviation of 10 Hz, and an AM modulation index of 0.25. The bandpass filter is Gaussian with a bandwidth of approximately 400 Hz. The outputs of the F/D stages are resolved but difficult to distinguish in this figure. (b) Close-up of a peak of modulation cycle showing the distinct curves of the F/D outputs (c) Close-up of a trough of the modulation cycle.

''wiggles'' in the curve follow the narrowband frequency modulation, while the gray scale of the curve varies with every cycle of the modulation as it follows the amplitude modulation (parts of the cycle are darker than others). For comparison purposes, Figs. 4(b)–(e) show the same signal analyzed with a conventional narrowband STFT, a wideband STFT, by using peak detection on the narrowband STFT, and by using peak detection on the wideband STFT. The additional detail that is provided by the fine structure spectrogram is clear.

5. Case of two-harmonic narrowband FM signal

To test the performance of the algorithm with a signal that has a more complex structure in the time-frequency plane than a simple sinusoid, a sinusoidal signal modulated by two harmonics is used. The carrier frequency of the FM signal is 1000 Hz, the modulation frequency and peak frequency deviation of the first harmonic are 45 and 20 Hz, respectively, and the modulation frequency and peak frequency deviation of the second harmonic are 90 and 10 Hz, respectively. Figure 5(a) shows the resulting FSS of the input signal, while Fig. 5(b) shows the calculated instantaneous frequency of the signal obtained by taking the derivative of the phase of the signal (Boashash, 1992a). Apart from a time shift in Fig. 5(a) due to the constant time delay in the FIR filters, the two figures are very similar. This demonstrates that the FSS can detect time-frequency fine structure within a single modulation cycle.

6. Case of linear chirp signal

Figure 6 shows the FSS of a chirp that sweeps from 1000 to 1500 Hz in 0.1 s. This is a frequency rate of change that is similar to that sometimes found in the glides that accompany vowel-consonant transitions in speech (see, for example, the second formant of the first three syllables in Fig. 9 below). Apart from the constant time shift due to the FIR filters, the fine structure spectrogram tracks this chirp quite well.

D. A Fourier transform representation

Anderson (1984) showed that a single F/D stage centered on ω_c consisting of a bandpass filter with an impulse response of $h_1(n)\cos\omega_c n$, a square-law device, and a low-pass filter with an impulse response of $h_2(n)$, is equal to the magnitude-squared of the STFT evaluated at ω_c:

$$|\mathrm{STFT}(\omega_c,n)|^2=\left|\sum_{m=-\infty}^{m=+\infty} h(n-m)x(m)e^{-j\omega_c m}\right|^2 \qquad (18)$$

if $h_1(n)=h_2(n)=h(n)$, and if $h(n)$ is the impulse response of the ideal low-pass filter.

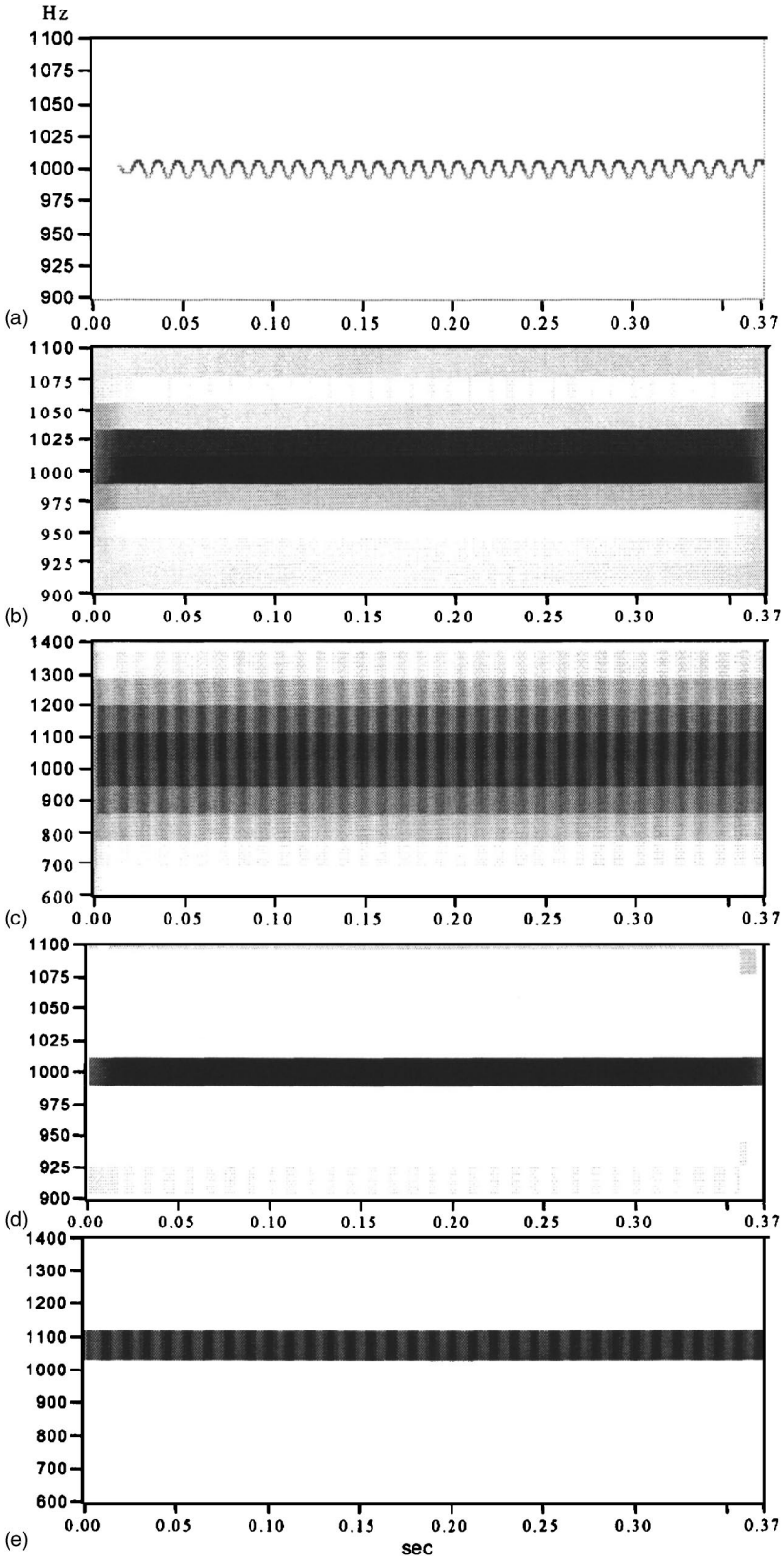


FIG. 4. (a) Fine structure spectrogram of the AM/NBFM signal described in Fig. 3 obtained with a direct computer implementation (bandpass filter bandwidth 200 Hz, low-pass filter cutoff frequency 100 Hz, filter separation 1 Hz, probing interval 0.045 ms). (b) Narrowband STFT of the same signal (46.4-ms 1024-point Hanning window shifted by 1.45 ms corresponding to a window overlap of 96.9%). (c) Wideband STFT of the same signal (5.8-ms 128-point Hanning window zero padded to 256 points and shifted by 1.45 ms corresponding to a window overlap of 75%). (d) With peak detection on the narrowband STFT. (e) With peak detection on the wideband STFT. Note the change in the *y* axis for the wideband STFT.

With realizable filters, the particular choice of $h(n)$, $h_1(n)$, and $h_2(n)$ determines how much the time-varying low-frequency envelope of the bandlimited signal is contaminated by higher frequency components, and so a single F/D stage and the magnitude-squared STFT are only approximately equal. Nevertheless, this equivalence formula says that the fine structure spectrogram can be implemented by calculating the magnitude-squared STFT evaluated at ω_c, if the bandpass filter bandwidth is constrained to be twice that of the low-pass filter. The potential advantage of this is that it allows the use of the wealth of theory and computational algorithms available for the STFT. But this equiva-

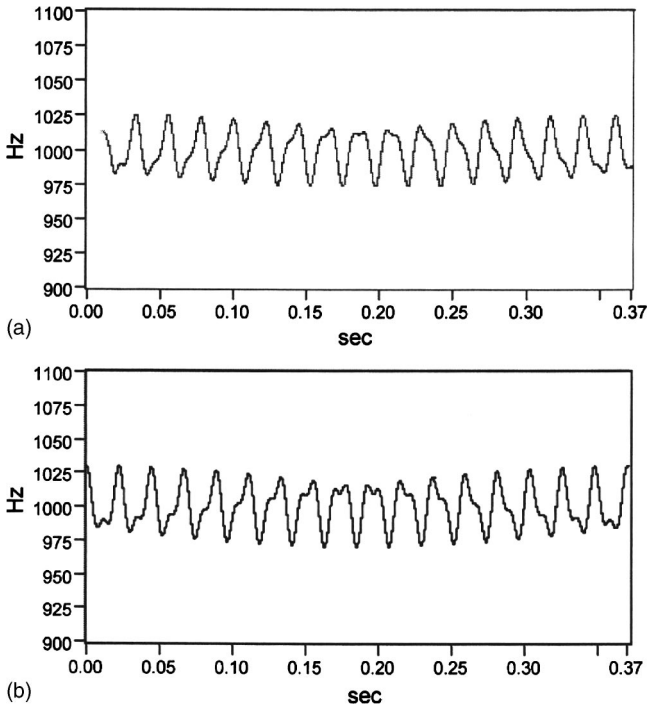


FIG. 5. (a) Fine structure spectrogram of a two-harmonic narrowband FM signal (bandpass filter bandwidth 400 Hz, low-pass filter cutoff frequency 150 Hz, filter separation 1 Hz, probing interval 0.045 ms). The carrier frequency of the FM signal is 1000 Hz, the modulation frequency and peak frequency deviation of the first harmonic are 45 and 20 Hz, respectively, and the modulation frequency and peak frequency deviation of the second harmonic are 90 and 10 Hz, respectively. (b) The derivative of the phase of the input signal, which estimates the instantaneous frequency.

lency raises the question of how the FSS is different from the conventional STFT, if it can be implemented with the magnitude-squared STFT. One difference is the ability to vary the bandwidths of bandpass and low-pass filters independently in the filterbank implementation of the FSS. Another difference is that if the STFT is used with the FSS, the STFT is calculated at many more frequency points ω_c than are used in the conventional STFT, and this corresponds to having a very large number of overlapping filters in the filterbank implementation of the algorithm (Sec. II A). Conventional spectrograms usually calculate the STFT either at the bin frequencies provided by the discrete Fourier transform or at a few additional intermediate frequencies between the bins. In fact, a research goal has been to determine the minimum allowable sampling rate along the time and frequency

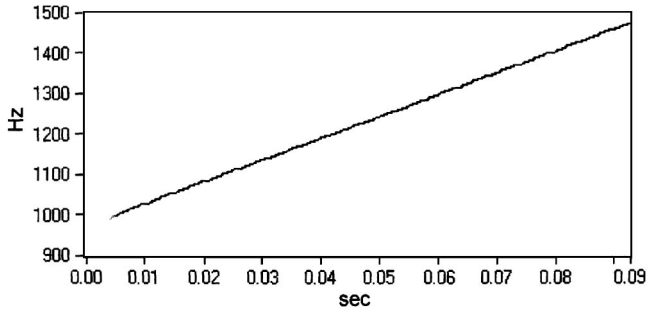


FIG. 6. Fine structure spectrogram of a linear chirp. The signal sweeps from 1000 to 1500 Hz in 0.1 s (bandpass filter bandwidth 400 Hz, low-pass filter cutoff frequency 200 Hz, filter separation 1 Hz, probing interval 0.045 ms).

dimensions that would result in no loss of information and so allow accurate signal reconstruction, often in the interest of computational efficiency (Rabiner and Schafer, 1978; Sorensen and Burrus, 1988; Oppenheim *et al.*, 1999).

Delprat *et al.* (1992) and Carmona *et al.* (1997, 1999) have studied the detection of local peaks in time-frequency representations, which they call ''ridge extraction.'' They proved that the ridges, in any time-frequency representation, can be used to estimate the instantaneous frequency and amplitude of components in the time-frequency plane. However, peak picking on the conventional spectrogram has been found to give poor results (Boashash, 1992b). The FSS, implemented using an ''oversampled'' STFT followed by a peak detector, allows demodulation of fine structure AM and FM components in the signal, which is not possible with the conventional implementation of the STFT. This oversampling can be performed by an evaluation of the STFT at many frequencies ω_c, or it can be achieved by zero-padding the analyzed segment with a very large number of zeros. In addition, the windowed segments that are used need to be short (which corresponds to having fairly wideband filters in the filterbank implementation of the algorithm), and the segments must be shifted by a small time interval to closely track the AM and FM modulations.

E. Error evaluation

1. Single AM/NBFM component

To evaluate the performance of the FSS as a function of the various parameters of modulated signals and of the FSS system itself, a sinusoidal signal that is jointly modulated in amplitude and frequency (AM/NBFM) at the same modulation frequency is used as a test signal and the FSS is implemented using a filterbank. Unless otherwise stated, the prototype test signal has a carrier frequency of 1000 Hz, a modulation frequency of 150 Hz, a peak frequency deviation of 10 Hz, and an AM index of 0.25. The prototype FSS system is implemented using bandpass filters that have a bandwidth of 400 Hz and a separation of 1 Hz, using low-pass filters with a bandwidth of 150 Hz, and probing the outputs of the F/D's every 0.045 ms. In each evaluation, one or more parameters of the test signal and the FSS system is varied. The root mean-squared error (RMSE) in Hz between the instantaneous frequency estimated by the FSS and the true instantaneous frequency is reported. The RMSE is evaluated after the initial transients due to the filter response have died down. In almost all cases, the FSS detected a signal modulated at the correct frequency, and the reported error resulted from differences in the frequency deviation. For the test signal with a peak deviation of 10 Hz, when the FSS detects the signal and the estimated frequency deviation is smaller than the true frequency deviation, the maximum possible RMSE is around 7 Hz. On the other hand, when the estimated frequency deviation exceeds the true frequency deviation, then the RMSE may be larger. Also, the instantaneous amplitude showed error trends similar to the error trends reported for the instantaneous frequency.

Figure 7(a) examines the effect of varying the bandpass filter bandwidths $\omega_{\mathrm{bpf}}(\omega_{\mathrm{bpf}} \geqslant 2\omega_m)$, for two low-pass cutoff

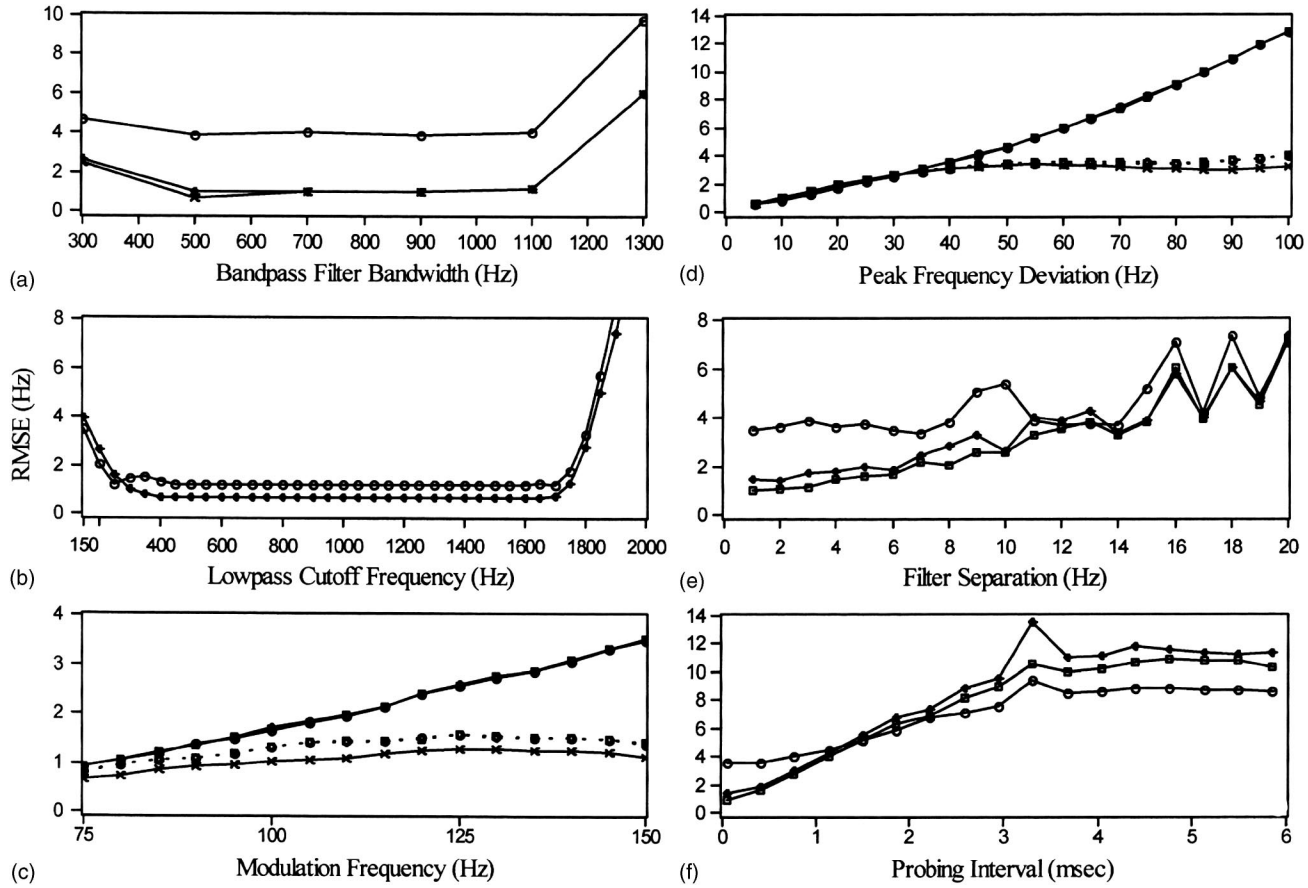


FIG. 7. RMSE (Hz) between FSS-estimated and true instantaneous frequency for an AM/NBFM signal. See text for a description of the parameters of the test signal and the FSS analysis system. The points designated by × were obtained using the Hilbert transform in place of rectification+low-pass filtering. (a) As a function of bandpass filter bandwidth (low-pass cutoff frequency: ○=150 Hz, ◇=300 Hz). (b) As a function of low-pass cutoff frequency (bandpass filter bandwidth: ○=400 Hz, ◇=800 Hz). (c) As a function of modulation frequency (low-pass cutoff frequency: solid line=150 Hz, dashed line=300 Hz; carrier frequency: ○=600 Hz, ◇=1000 Hz, □=2500 Hz, ×=1000 Hz). Curves at the three carrier frequencies practically coincide. (d) As a function of peak frequency deviation (low-pass cutoff frequency: solid line=200 Hz, dashed line=400 Hz; carrier frequency: ○=600 Hz, ◇=1000 Hz, □=2500 Hz, ×=1000 Hz). Curves at the three carrier frequencies practically coincide. (e) As a function of filter separation (low-pass cutoff frequency, bandpass filter bandwidth: ○=150, 400 Hz, ◇=300, 400 Hz, □=300, 800 Hz). (f) As a function of probing interval in ms (low-pass cutoff frequency, bandpass filter bandwidth: ○=150, 400 Hz, ◇=300, 400 Hz □=300, 800 Hz).

frequencies ω_{lpf} (150 and 300 Hz). As ω_{bpf} increases beyond $2\omega_m$, there is an initial reduction in the estimation error as the sidebands of the signal are better represented at the output of the bandpass filter [Eqs. (14) and (15)]. The estimation error then remains stable beyond approximately $4\omega_m$ until the ω_{bpf} is so large ($\gg 2\omega_m$) that there is a loss of frequency localization across of F/D outputs and hence poor time-frequency responsiveness. When a wider low-pass filter is used (ω_{lpf} of 300 Hz vs. 150 Hz), the estimation error follows a similar trend but is consistently smaller. The effect of varying the low-pass filter cutoff frequency ω_{lpf} ($\omega_{\mathrm{lpf}} \geqslant \omega_m$) is examined in Fig. 7(b) for two bandpass filter bandwidths (400 and 800 Hz). There is a sharp drop in the estimation error for $\omega_m \leqslant \omega_{\mathrm{lpf}} \leqslant$(approximately)$2\omega_{\mathrm{m}}$. As the low-pass cutoff frequency increases beyond around $2\omega_m$, the inclusion of components at $2\omega_m$, $3\omega_m$, and $4\omega_m$ [Eq. (16)] causes little change in the performance as they are highly diminished relative to the component at ω_m. However, further increasing ω_{lpf} to include components at much higher frequencies in the output of the low-pass filter results in a dramatic increase in the estimation error. Figure 7(c) examines the effect of varying the modulation frequency, for three carrier frequencies (600, 1000, and 2500 Hz), for two low-pass cutoff frequencies ω_{lpf} (150 and 300 Hz). The estimation error is independent of the carrier frequency. For the case of $\omega_{\mathrm{lpf}}=150\,\mathrm{Hz}$, as the modulation frequency varies from $0.5\omega_{\mathrm{lpf}}$ to ω_{lpf}, there is an increase in the estimation error as the bandpass and low-pass filters become less capable of tracking the instantaneous changes in the signal. However, for the case of $\omega_{\mathrm{lpf}}=300\,\mathrm{Hz}$, as the modulation frequency varies from $0.25\omega_{\mathrm{lpf}}$ to $0.5\omega_{\mathrm{lpf}}$, there is little change in the estimation error. Both these effects were previously seen for the signal in Fig. 7(b) for $\omega_m \leqslant \omega_{\mathrm{lpf}} \leqslant 4\omega_m$. Figure 7(d) shows the effect of increasing the peak frequency deviation of the signal for three carrier frequencies (600, 1000, and 2500 Hz), a modulation frequency of 100 Hz, and for two low-pass cutoff frequencies (200 and 400 Hz). The estimation error is independent of carrier frequency, but better performance is achieved with the wider low-pass filter, particularly at higher peak frequency deviations. Figure 7(e) shows the effect of increasing the separation between the filters, for three bandpass filter bandwidth and low-pass filter cutoff frequency pairs: (150, 400 Hz), (300, 400 Hz), and (300, 800 Hz). With larger filter separations, the system

has more difficulty in tracking the structure of the instantaneous frequency. The best performance is achieved at small filter separations and the widest filter bandwidths ($\omega_{bpf} > 4\omega_m$ and $\omega_{lpf} = 2\omega_m$). Finally, Fig. 7(f) examines the effect of progressively increasing the probing interval at the output of the F/D's for three bandpass filter bandwidth and low-pass filter cutoff frequency pairs: (150, 400 Hz), (300, 400 Hz), and (300, 800 Hz), without decreasing the sampling rate of the signal. The estimated instantaneous frequency is assumed to remain constant for the duration of the probing interval, making this analogous to assuming constant spectral content within a windowed segment in quasi-stationary analysis. The estimation errors are lowest at small probing intervals, where the best performance is again achieved with the widest filter bandwidths ($\omega_{bpf} > 4\omega_m$ and $\omega_{lpf} = 2\omega_m$). In summary, the results of these tests point to an operating range for the bandpass filter bandwidths of $2\omega_m \leqslant \omega_{bpf} \leqslant \omega_{max}$ ($\omega_{max} \gg \omega_m$), with an optimum range of $4\omega_m \leqslant \omega_{bpf} \leqslant \omega_{max}$, and an operating range for the low-pass filter cutoff frequencies of $\omega_m \leqslant \omega_{lpf} \leqslant \omega_{max}$ ($\omega_{max} \gg \omega_m$), with an optimum range of $2\omega_m \leqslant \omega_{lpf} \leqslant \omega_{max}$. Moreover, if the filters have gentle roll-offs, then bandwidths outside these ranges may be used, but at the cost of degradation in estimation performance. The results also indicate that smaller estimation errors in tracking the instantaneous frequency are possible only if small filter separations and small probing intervals are used.

2. Comparison with the Hilbert transform in place of rectification and smoothing

The Hilbert transform (HT), used to construct the analytic signal, is a standard approach for estimating the instantaneous frequency and envelope of a modulated signal (Potamianos and Maragos, 1994). To track the frequency and amplitude of a speech resonance, the resonance is typically first isolated using a fixed bandpass filter. The problem with this approach is that the center frequency of a speech formant may vary considerably during an utterance (for example, see the second formant in the word /oily/ in Fig. 9). And even if a formant has a fairly steady center frequency, such as the third formant in Fig. 9, the output of a bandpass filter that is centered near this formant may switch to reporting the instantaneous frequency and amplitude of the—typically stronger—second formant when the center frequency of the second formant rises to approach the center frequency of the filter.

An alternative approach for tracking speech resonances that uses the HT is to replace the rectification and low-pass filtering in the FSS with a stage that outputs the magnitude of the analytic signal. The performance of this approach was evaluated using the same test signal discussed in the previous section. Figure 7(a) shows the RMSE obtained using the HT +FSS as a function of the bandpass filter bandwidth, for a modulation frequency of 150 Hz. Figure 7(c) shows the performance of the HT+FSS as a function of modulation frequency, for a carrier frequency of 1000 Hz. Figure 7(d) shows the performance of the HT+FSS with increasing peak frequency deviation, for a carrier frequency of 1000 Hz. In Figs. 7(a), (c), and (d), the curves obtained using rectification and low-pass filtering with wider low-pass filters are closer to the RMSE obtained using the HT than those obtained with narrower low-pass filters.

3. Multiple AM/NBFM components

To evaluate the performance of the FSS when multiple modulated components are present, a test signal consisting of two AM/NBFM components is used. The carrier frequency of one component is fixed at 1000 Hz, while the carrier frequency of the second component ranges from 500 to 800 Hz. The peak frequency deviation for both components is 10 Hz, the AM modulation index is 0.25, and unless otherwise stated the modulation frequency is 150 Hz. The bandpass filters in the FSS system are separated by 1 Hz, the F/D outputs are probed every 0.045 ms, and unless otherwise stated the bandwidth of the bandpass filters is 400 Hz and the low-pass cutoff frequency is 300 Hz. The RMSE between the estimated and true instantaneous frequencies is averaged over the two components. In most cases, as the two components got closer, the performance of the FSS degraded at first gradually and then reached a transition region where it degraded drastically. However, even when the estimated error became very large, it was often still possible to see two distinct—though highly distorted—components. In this case, the FSS would be useful when there is more interest in investigating the presence of different time-varying modulated components in a signal than in the accuracy of the frequency estimations.

Figure 8(a) shows the RMSE for two modulation frequencies (50 and 150 Hz) as a function of the carrier separation frequency. With larger carrier separations (above approximately 350 Hz), a smaller estimation error is obtained with the lower modulation frequency of 50 Hz. This is expected because the larger spread of the spectral components at higher modulation frequencies would result in more interaction between the two modulated components. The input signal for the case of a modulation frequency of 50 Hz is shown in Fig. 8(d), while Figs. 8(e) and (f) show the FSS that is obtained with carrier separations of 425 and 250 Hz, respectively. With the smaller carrier separation, the two modulated components highly interact, though it is arguably possible to distinguish components that are centered on the carrier frequencies. It is possible to obtain better results at this carrier separation, with much less interaction between the two components, if narrower bandpass filters are used. Figure 8(b) shows the RMSE for three bandpass filter bandwidths (200, 300, and 400 Hz) as a function of carrier separation. At small carrier separations, smaller bandpass filter bandwidths result in a smaller RMSE. Note that even though the 200-Hz bandpass filter falls outside the optimum operating range for the individual modulated components in this signal, it was still possible to obtain an estimate of the instantaneous frequencies of the components because of the gentle roll-offs of the filters (Sec. II E 1). Figure 8(c) shows the RMSE for two low-pass cutoff frequencies (150 and 300 Hz) as a function of carrier separation. For the larger bandwidth of the low-pass filter, the performance degrades drastically at larger carrier separations. As the 300-Hz low-pass filter has a gentle roll-off, it admits more high-frequency

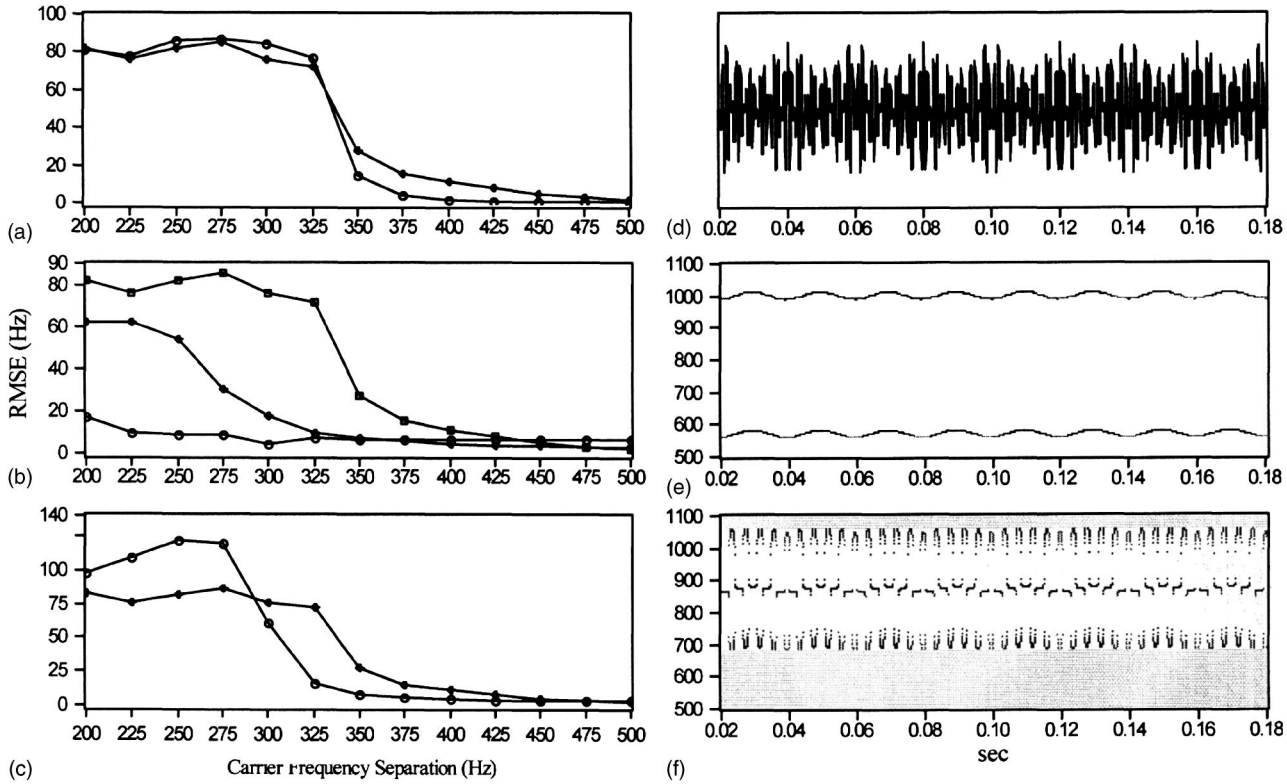


FIG. 8. RMSE (Hz) between FSS-estimated and true instantaneous frequency for a signal consisting of two AM/NBFM components, as a function of the separation between the two components. See text for a description of the parameters of the test signal and the FSS analysis system. (a) Modulation frequency: ○=50 Hz, ◇=150 Hz. (b) Bandpass filter bandwidth: ○=200 Hz, ◇=300 Hz, □=400 Hz. (c) Low-pass cutoff frequency: ○=150 Hz, ◇=300 Hz. (d) Time domain signal in (a) with a modulation frequency of 50 Hz. (e) FSS obtained in (a) with a carrier separation of 425 Hz. (f) FSS obtained in (a) with a carrier separation of 250 Hz.

products of interactions between the two components. In summary, larger bandpass and low-pass filter bandwidths (within their operating ranges), which are often beneficial for single component estimation, may increase the estimation error if two components are close in frequency. In the case of multiple components, it is therefore preferable to use the narrowest filters, within their operating ranges.

III. APPLICATION TO SPEECH

Rao and Kumaresan have recently developed a model-based method for decomposing the speech signal into modulated components (Kumaresan and Rao, 1999; Rao and Kumaresan, 2000), which represent the formants in the case of voiced speech, using adaptive filters whose center frequencies closely follow the resonance frequencies. Figure 9(a) shows the time-domain signal for the sample ''...an oily rag like that'' spoken by a male. Figures 9(b), (e), and (f) show the conventional narrowband STFT for this signal, the result of peak picking on this narrowband STFT, and the result of peak picking on the conventional wideband STFT. In these figures, the formants are evident but without much fine structure detail. Figure 9(c) is a reprint of the result obtained by Rao and Kumaresan (2000). The dark lines in the graph are estimates of the instantaneous frequencies of the modulated components, while amplitude variations are shown in a separate graph (not reproduced here). These lines are superimposed on a gray scale background consisting of the wideband STFT of the sample, and in which the vertical striations indicate the pitch cycles. Much finer detail is visible in the formants [compared to Figs. 9(b), (e), and (f)]. In particular, frequency modulations occur with every pitch cycle and, due to time compression in the figure, appear as periodic ''spikes'' in the formants. Figure 9(d) shows the result obtained with the fine structure spectrogram, which [unlike Fig. 9(c)] incorporates the amplitude information in the gray scale.

Overall, the FSS [Fig. 9(d)] shows details in the modulations that are similar to those seen in Fig. 9(c), but with some of the higher frequency formants hidden due to the range of the gray scale. The FSS, however, shows some time-frequency components not evident in Fig. 9(b), (c), (e) or (f). For example, in Fig. 9(d) there are components at frequencies below the first formant, a branching component at around 700 Hz between 0.2 and 0.45 s, and another one at around 1000 Hz between 0 and 0.1 s. That the additional components are not artifacts of the signal processing is supported by the fact that the FSS has a simple physical interpretation: the bandpass filters respond to input energy in their frequency band. To investigate this question further, the branching component in Fig. 9(d) (at around 700 Hz and between 0.25 and 0.45 s) was selected because it appears to have an approximately steady carrier frequency for a fairly long period, and appears to be distinctively separated from the first and second formants.

The time-domain signal at the output of a bandpass filter centered at 855 Hz with a bandwidth of approximately 200

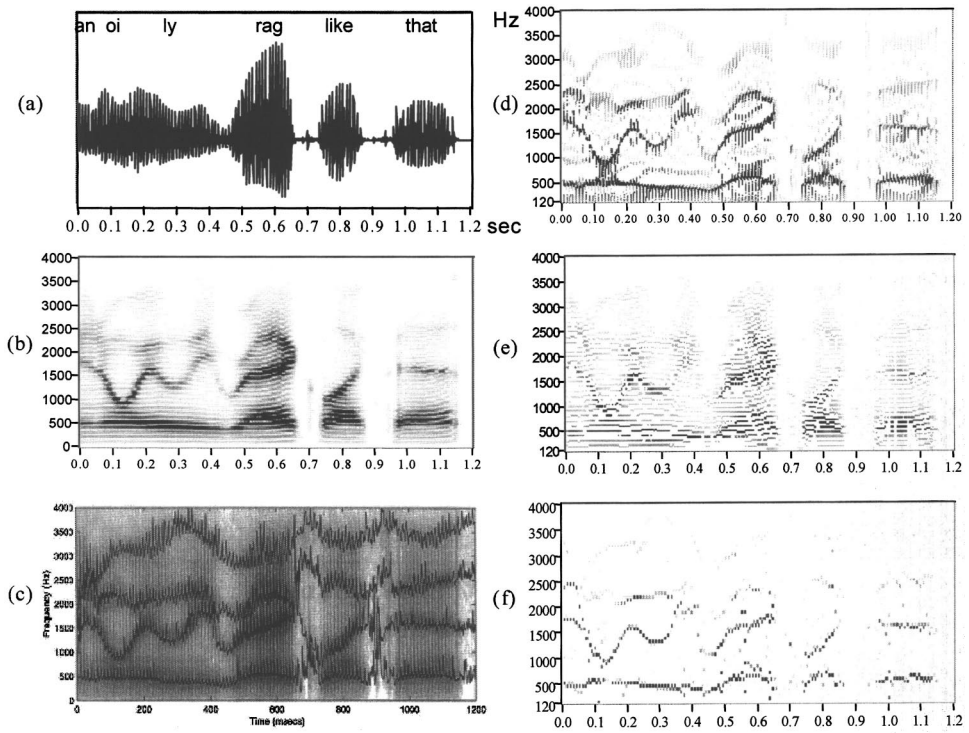


FIG. 9. Analysis of "...an oily rag like that" spoken by a male (from the TIMIT database). (a) Time-domain signal. (b) Narrowband STFT (32-ms Hanning window). (c) Modulated components obtained with the method of Rao and Kumaresan (2000) with wideband STFT as background. © 2000 IEEE. Reprinted with permission. (d) Fine structure spectrogram (4-ms 64-point Hanning window zero padded to 4096 points, with window separation of 0.0625 ms). (e) With peak detection on the narrowband STFT. (f) With peak detection on the wideband STFT (4-ms Hanning window).

Hz was extracted (a similar result is obtained if the bandpass filter has a bandwidth of 100 Hz). The center frequency of the bandpass filter was chosen to be higher than the apparent carrier frequency of the branching component to reduce contamination by the higher amplitude first formant at around 400 Hz. Between 0.35 and 0.4 s there were 34 cycles in this signal, corresponding to a carrier frequency of around 680 Hz. The time locations of the peaks of these cycles were determined with the same peak detection algorithm as that used in the FSS, and the reciprocals of the interpeak intervals were calculated in order to provide an estimate of the instantaneous frequency (Fig. 10). This approach to estimate the instantaneous frequency is similar to the zero-crossing method sometimes used for that purpose (Boashash, 1992b). To estimate the modulation parameters of the signal in Fig. 10, the first three modulations, which are the most stable, are used. The mean frequency, which is an estimate of the carrier, is 719 Hz. The average duration of the modulation cycle is 10.1 ms, giving a modulation frequency of 99 Hz. The frequency deviation ranges from an average of 457 to 924 Hz. This analysis supports the existence of the branching component shown in Fig. 9(d). There is, however, some discrepancy in that the component in that figure appears to span a smaller frequency range than the analysis above indicated. A possible explanation is that the dip around 400 Hz in the instantaneous frequency of Fig. 10 may result from the high amplitude component in the speech sample at around 400 Hz, and the branching component seen in the FSS may be reflected in the structure of the instantaneous frequency between around 600 and 900 Hz in this figure. The speech sample was also analyzed using other time-frequency representations. The Choi-Williams, pseudo Wigner-Ville, and cone-shaped kernel distributions (Cohen, 1995) did not show the additional branching component investigated above. However, one distribution, the Gabor spectrogram (Qian and Chen, 1999), does show a region of energy at the location of the branching component seen in the FSS (Fig. 11).

It should be pointed out that these additional components cannot be artifacts due to cross terms in the equations

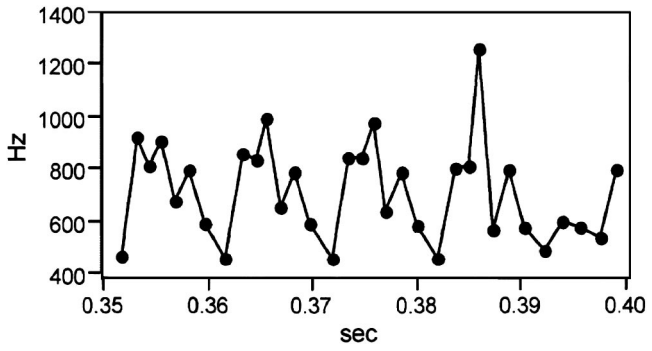


FIG. 10. Reciprocals of the interpeak intervals in the time-domain output (between 0.35 and 0.4 s) of a bandpass filter centered at 855 Hz, for the signal of Fig. 9. The bandwidth of the filter is approximately 200 Hz. The reciprocals of the interpeak intervals provide an estimate of the instantaneous frequency of the signal.

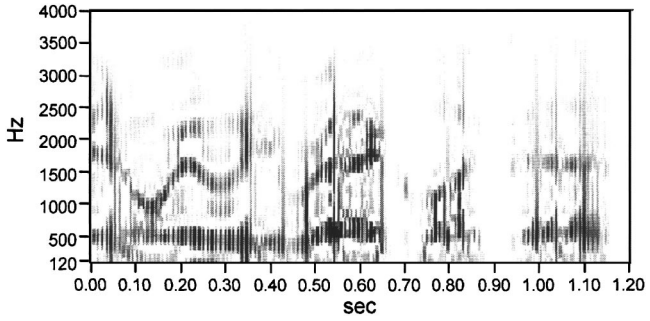


FIG. 11. Gabor spectrogram (8-ms equivalent window length) of the sentence analyzed in Fig. 9.

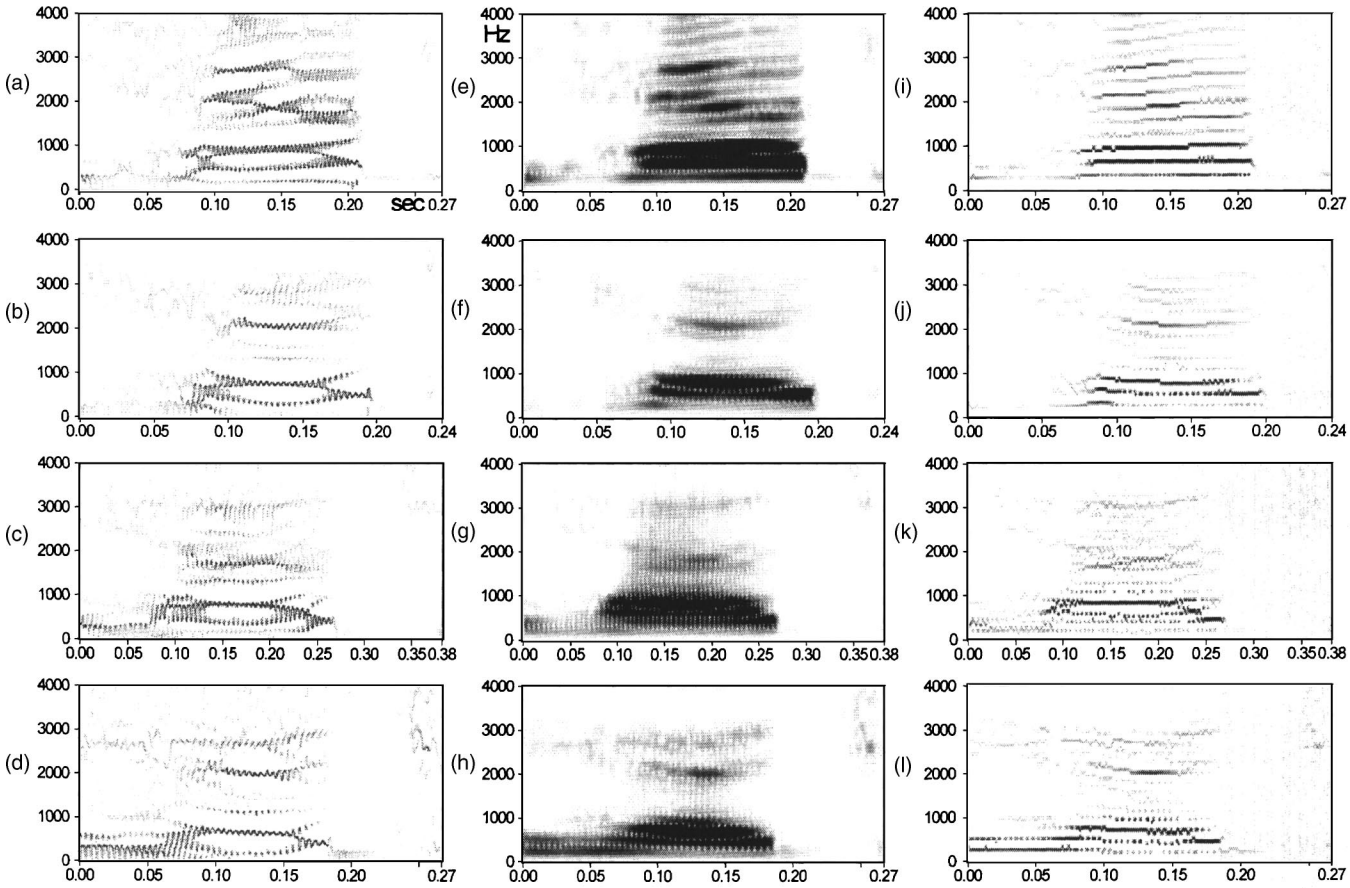


FIG. 12. (a)–(d) FSS (4-ms 64-point Hanning window zero padded to 4096 points, with window separation of 0.0625 ms) of the word /had/ spoken by four females (test/dr1/faks0/sa1, test/dr4/fadg0/sa1, train/dr2/fajw0/sa1, and test/dr5/fcal1/sa1 in the TIMIT database). (e)–(h) Wideband spectrogram (8-ms Hanning window) of the same words. (i)–(l) With peak detection on the wideband spectrogram.

used to calculate the FSS, and which are often seen in bilinear time-frequency representations. When the FSS is implemented using the magnitude-squared STFT (Sec. II D), cross terms due to the nonlinearity coincide with the auto terms and only affect the relative heights of the peaks (Kootsookos *et al.*, 1992; Cohen, 1995). Moreover, as these components are often as high as the second formant and sometimes in the range of only a few decibels below the first formant, they cannot be side lobes of the Hanning window, the highest of which are 31 dB below the primary lobe (Oppenheim *et al.*, 1999). It is also true that if two different modulated components are very close in frequency, they may interact to produce time-frequency structure in the FSS at an intermediate frequency between them [e.g., see Fig. 8(f)]. However, for the duration of the branching component investigated above, the first and second formants are separated by between 900 and 1400 Hz (as determined using the various methods used in Fig. 9). In simulations with two steady modulated components that have rates of modulation that roughly mimic the first two formants, and with the FSS obtained using the same parameters used for Fig. 9(d), there is no interaction between the two components unless their separation in frequency is made much smaller.

We speculate that conventional methods of speech processing, which are almost always based on the source-filter model of speech production, may miss the extra components because they assume that the individual lines of the spectrum of speech are simply harmonics of the source. In fact, as this study and other studies that decompose speech into modulated components show, the lines in the spectrum may be (at least partially) more properly viewed as sidebands of AM/FM components. And if the modulation frequency is equal (or approximately equal) to the fundamental frequency of speech, then the sidebands would be positioned in the same bin as the ''true'' harmonics of the fundamental frequency, making it difficult to distinguish them from each other. The FSS, in contrast, may combine the sidebands of AM/FM modulation into distinct time-frequency modulated components. Other methods for decomposing speech into AM/FM components that are model-based do not appear to show these extra components (Maragos *et al.*, 1993; Kumaresan and Rao, 1999), possibly because they make assumptions about the number of modulated components in the signal.

Some other examples of the FSS of speech samples are shown in Fig. 12, for the word /had/, from the sentence ''she had your dark suit in greasy wash water all year'' in the TIMIT database (test/dr1/faks0/sa1), spoken by four females. In the FSS of all four utterances [Figs. 12(a)–(d)], there is a prominent component below the first formant which, for /æ/ and adult females, is located on average at 860 Hz in American English (Peterson and Barney, 1952). This low-

frequency component is supported by the distinct band of low-frequency energy seen in the wideband spectrogram [Figs. 12(e)–(h)]. Ladefoged (2003) has noted the presence of significant spectral energy below the first formant in some speakers. Above the first formant, the FSS of each of the four samples has a distinctive structure, sometimes having a much more complex time-frequency structure than the traditional picture of formants as single moving tracks in the time-frequency plane [e.g., Fig. 12(a)]. Figures 12(i)–(l) show the results of peak picking on the wideband spectrograms. These figures show some fine structure but with less detail than the FSS in Figs. 12(a)–(d).

IV. DISCUSSION

This paper presents a method that may be interpreted as an enhanced realization of the conventional spectrogram for the analysis of signals whose spectral content varies with time, and of speech in particular. The main feature of the method, referred to as the fine structure spectrogram (FSS), is the very large number of overlapping filter/detector (F/D) stages, followed by a peak detector. When viewed as a bank of F/D's, the FSS achieves good time-frequency responsiveness by having fairly wideband filters, a large number of filters separated by a very fine amount, and probing the outputs of the F/D's very frequently. The FSS may also be implemented using peak picking on a highly ''oversampled'' magnitude-squared STFT. While in the more generalized FSS the bandwidths of the bandpass and low-pass filters can be varied independently, if the bandwidth of the bandpass filter is constrained to be twice that of the low-pass filter, there is no practical difference analytically between the filterbank-based FSS and peak picking on a very high resolution spectrogram, provided (1) the windowed segments are short, (2) they are very highly overlapping, (3) the STFT is calculated at a very large number of frequency points, and (4) the impulse response of the bandpass filters is a frequency-shifted version of the impulse response of the low-pass filters. If these conditions are not satisfied, analysis of the estimation error indicates that peak picking on the spectrogram would not be effective in tracking the fine structure of speech modulations. The spectrogram is typically implemented with partially overlapping windows, or with a few partially overlapping filters, as it has been argued that all the information is preserved if the spectrogram is implemented as such (Rabiner and Schafer, 1978). Perhaps, for this reason, peak picking on the conventional spectrogram has not been previously shown to effectively extract the fine structure of frequency and amplitude modulations in the resonances of speech with high time-frequency resolution.

We emphasize, however, that one should be very careful about defining what is meant by time-frequency resolution, as it is important to distinguish between the ability to resolve simultaneous frequency components, and the ability to track the frequency of a modulated component. With several types of modulated signals, the FSS can effectively track the changing frequency and amplitude of components in the time-frequency plane, and so show details in the modulations that are not visible using quasi-stationary methods. It cannot, however, resolve components that are very closely spaced in frequency because the bandwidth of the filters controls the ability to resolve simultaneous frequency components. Nelson (2001) describes the ability to provide accurate time and frequency estimates of modulated components, beating the uncertainty limit of the STFT, as ''super-resolution.'' However, it may be more appropriate to call it high time-frequency ''discrimination,'' following hearing science terminology where it is known that the auditory system is much better at *discriminating* between two sequential tones than at *resolving* two simultaneous tones (Houtsma, 1995). If modulated components approach each other in the time-frequency plane, large estimation errors in the instantaneous frequency may result. However, in that case, and depending on the bandwidth of the bandpass and low-pass filters, the FSS will often still detect two distinct components.

The FSS also detects additional components in speech that may be missed by other methods. The spectral energy of these components is usually significant relative to the known formants. We believe that these component are often real, because the FSS has a simple physical interpretation: the bandpass filters respond to input energy in their frequency band. This is supported by a detailed investigation into one of the extra components (Sec. III). As to the physical meaning of the extra components and the glottal cycle-by-cycle modulations seen in the FSS, there is now strong theoretical and experimental evidence for the existence of significant nonlinear aerodynamic phenomena in speech that cannot be accounted for by the commonly used linear model of speech production (Thomas, 1986; Teager and Teager, 1989; Maragos *et al.*, 1993, 2002). Maragos *et al.* (2002) divide these nonlinear phenomena into two types: modulations and turbulence. Amplitude and frequency modulations can be generated within a glottal cycle due to airflow separation, in which the air flowing through the vocal tract oscillates between its walls. In addition, vortices, which are easily generated in the vocal tract, modulate the energy of the air jet, and may result in instabilities leading to turbulent air flow. How these nonlinear mechanisms would result in extra ''branching components'' in the FSS is not clear, but nonlinear interactions between the source and the filter (Quatieri, 2002) could plausibly generate resonances at frequencies other than those predicted by the linear source-filter theory. Wind instruments may provide a useful analogy, where nonlinear interaction between the driving mechanism and the resonant tube typically produces complex sounds with inharmonic components, bifurcations, and other nonlinear dynamic phenomena (Maganza *et al.*, 1986; Fletcher, 1990, 1998). Similar nonlinear phenomena have been observed in some types of speech vocalizations, particularly in patients with voice disorders (Herzel *et al.*, 1998). To the aforementioned nonlinear mechanisms, we can add a linear mechanism that results in modulations between glottal cycles. Passing a series of pulses (representing the glottal source) through a linear bandpass filter (representing the vocal tract) produces cycle-by-cycle amplitude and frequency variations (Flanagan, 1980; Potamianos and Maragos, 1994).

With the additional detail in the time-frequency plane and the extra components that are visible, the FSS provides a picture of speech that is somewhat different from the tradi-

tional view of the resonances of speech. In the conventional view, and at least with voiced speech, the resonances of the vocal tract are thought of as more or less single tracks that glide smoothly in the time-frequency plane. In contrast, the FSS often shows a much more complex pattern for speech in the time-frequency plane. We therefore speculate that the FSS may be useful in identifying the time-frequency components of speech that are essential for producing natural sounding synthetic speech, or in pattern recognition applications where the task is to distinguish similar sounding speakers or phonemes. The utility of the AM/FM structure in speech has been investigated by Jankowski *et al.* (1996) in a speaker identification application, by Tolba and O'Shaughnessy (1998) in a speech recognition application, and by Hertich and Ackermann (1999) in a speech synthesis application. The practical utility of detecting extra components, however, remains to be seen, and would require the analysis of a larger set of utterances from more speakers than presented in this paper. In pattern recognition applications, a practical problem is that the FSS results in a very large 2-D data set. To overcome this, Rao and Kumaresan (2000) suggested downsampling the detected modulations to construct feature vectors. Machine recognition applications, of course, also require a degree of noise robustness, so in cases of substantial noise, the data may need to be denoised as an initial step.

The ability of the FSS to track the fine structure of AM and FM modulations in speech may also help explain the ability of the auditory system to perform these tasks. Quatieri *et al.* (1997) maintain that current models of auditory processing that use a filterbank cannot adequately explain the ability of the auditory system to track frequency modulations. They argue that the mechanism for frequency demodulation by the ear is not known, but discuss a number of candidates, one of which is the possible tracking of the frequencies at which the local maxima in the neural firing rate occur. However, they do not develop this approach, which bears a strong resemblance to the FSS, any further. Therefore, as a matter of speculation, and notwithstanding the need to incorporate constant-Q filters that more closely resemble auditory tuning curves, the results in this paper suggest that the FSS may provide a framework to explain frequency and amplitude demodulation by the auditory system. This framework may be especially applicable at higher frequencies where the critical bands are broad, but not at low frequencies, where the narrower auditory filters may decompose formants or formant-like modulated components into harmonically related components.

ACKNOWLEDGMENTS

This work was supported by the National Research Council of Canada (NSERC) and an Ontario Graduate Scholarship. We would like to thank the anonymous reviewers for their suggestions and criticisms that helped improve this paper.

APPENDIX: DERIVATION FOR NBFM AND AM/NBFM SIGNALS

This appendix details the derivation of two results related to narrowband FM (NBFM) and combined AM/narrowband FM (AM/NBFM) signals. The derivations follow from three basic assumptions:

(1) The peak frequency deviation of the signals is small compared to the carrier frequency.
(2) Terms dependent on higher powers of m_{AM} and m_{FM} (and their cross-terms) can be ignored.
(3) The standard deviation of the Gaussian filters is large compared to the modulation frequency of the FM signal. That is,

$$\sigma^2 \gg \omega_m^2.$$

These assumptions follow from the choice of synthetic signals used to test the FSS. Assumption 1 is satisfied by the narrowband approximation and the observation that the carrier frequency is typically much greater than the modulation frequency. Regarding Assumption 2, since m_{AM} and m_{FM} are both less than one, terms in higher power of m_{AM} and m_{FM} (including any cross-terms, e.g., $m_{AM}m_{FM}$) can be ignored with little loss in numerical accuracy. Assumption 3 is satisfied since the bandwidth of the Gaussian filters must be wider than at least twice the modulation frequency in order to pass the signal. (Note that the filter bandwidth is linearly proportional to the standard deviation for a Gaussian filter.) Another way to evaluate the veracity of the assumptions is to compare the numerical calculations alongside the theoretical predictions.

1. Narrowband FM signal

Beginning with the output from the low-pass filter from Eq. (12),

$$(A_1^2+A_2^2+A_3^2)/2+A_3(A_2-A_1)\cos\omega_m t, \tag{A1}$$

the amplitudes of the respective components are given by

$$A_1=\tfrac{1}{2}m_{FM}A_0\exp[-(\omega_c-\omega_m-\mathrm{CF})^2/2\sigma^2],$$

$$A_2=\tfrac{1}{2}m_{FM}A_0\exp[-(\omega_c+\omega_m-\mathrm{CF})^2/2\sigma^2],$$

$$A_3=A_0\exp[-(\omega_c-\mathrm{CF})^2/2\sigma^2].$$

We differentiate Eq. (A1) with respect to the center frequency CF and set this expression equal to zero. The result is a complex expression that can be simplified by first making the substitution

$$\mathrm{CF}=\omega_c[1+\Delta(t)], \tag{A2}$$

where $\Delta(t)$ is a time-varying function. In Eq. (A2) we have simply reexpressed the center frequency (of maximal output) as some time-varying displacement about the carrier frequency. Next we assume that $|\Delta(t)|\ll 1$ (which follows from assumption 1) and take a first-order series expansion in Δ to obtain

$$\Delta = \frac{2m_{FM}\omega_m\sigma^2 \exp(-\omega_m^2/2\sigma^2)\cos(\omega_m t)}{\omega_c[2\sigma^2 + m_{FM}^2\sigma^2 \exp(-\omega_m^2/\sigma^2) - 2m_{FM}^2\omega_m^2 \exp(-\omega_m^2/\sigma^2)]}. \quad \text{(A3)}$$

This expression can be simplified further by ignoring terms of order m^2 and higher and by taking $\exp(-\omega_m^2/\sigma^2) \cong 1$ to obtain

$$\Delta = \frac{m_{FM}\omega_m \cos(\omega_m t)}{\omega_c} \quad \text{(A4)}$$

or

$$\mathrm{CF} = \omega_c + m_{FM}\omega_m \cos(\omega_m t). \quad \text{(A5)}$$

Equation (A5) is the derivative of the phase for the original FM signal and thus it shows that the FSS correctly demodulates the signal. With the test NBFM signal explored in Sec. II C 3, the absolute error between Eq. (A5) and the estimation result from the FSS is $<0.12\%$ for the duration of the modulation cycle. The two, therefore, are in good agreement.

2. Combined AM/narrowband FM signal

The derivation for an AM/NBFM signal proceeds in a similar manner. Starting with the output from the low-pass filter from Eq. (17),

$$(A_1^2 + A_2^2 + A_3^2 + A_4^2 + A_5^2)/2 + (A_1A_3 + A_2A_3 + A_1A_4 + A_2A_5)\cos\omega_m t, \quad \text{(A6)}$$

the amplitudes of the five components are given by

$$\begin{aligned} A_1 &= \tfrac{1}{2}(m_{AM} - m_{FM})A_0 \exp[-(\omega_c - \omega_m - \mathrm{CF})^2/2\sigma^2], \\ A_2 &= \tfrac{1}{2}(m_{AM} + m_{FM})A_0 \exp[-(\omega_c + \omega_m - \mathrm{CF})^2/2\sigma^2], \\ A_3 &= A_0 \exp[-(\omega_c - \mathrm{CF})^2/2\sigma^2], \\ A_4 &= -\tfrac{1}{4}m_{AM}m_{FM}A_0 \exp[-(\omega_c - 2\omega_m - \mathrm{CF})^2/2\sigma^2], \\ A_5 &= \tfrac{1}{4}m_{AM}m_{FM}A_0 \exp[-(\omega_c + 2\omega_m - \mathrm{CF})^2/2\sigma^2]. \end{aligned} \quad \text{(A7)}$$

Since both A_4 and A_5 are functions of higher powers of m_{AM} and m_{FM}, they are ignored. When A_4 and A_5 are set to zero, Eq. (A6) is identical to Eq. (A1) and therefore the equation for CF is identical to the one derived for the NBFM case. (note that the amplitudes in Eq. (A6) are signed, while they are unsigned in Eq. (A1))

Next we check to see if the AM component is correctly demodulated. Substituting Eq. (A5) into the reduced form of Eq. (A6) (i.e., with $A_4 = A_5 = 0$), we obtain

$$\begin{aligned} &\tfrac{1}{8}(m_{AM} - m_{FM})^2A_0^2C_1 + \tfrac{1}{8}(m_{AM} + m_{FM})^2A_0^2C_2 + \tfrac{1}{2}A_0^2C_3 \\ &+ \tfrac{1}{2}(m_{AM} - m_{FM})A_0^2C_1C_3\cos(\omega_m t) \\ &+ \tfrac{1}{2}(m_{AM} + m_{FM})A_0^2C_2C_3\cos(\omega_m t), \end{aligned} \quad \text{(A8)}$$

where

$$\begin{aligned} C_1 &= \exp[-(-\omega_m - \omega_m m_{FM}\cos(\omega_m t))^2/\sigma^2], \\ C_2 &= \exp[-(\omega_m - \omega_m m_{FM}\cos(\omega_m t))^2/\sigma^2], \\ C_3 &= \exp[-\omega_m^2 m_{FM}^2 \cos^2(\omega_m t)/2\sigma^2]. \end{aligned}$$

The first two terms in Eq. (A8) are dropped and all exponential terms are set equal to one. Thus we obtain for the amplitude of the filter of maximum output

$$\tfrac{1}{2}A_0^2 + \tfrac{1}{2}(m_{AM} - m_{FM})A_0^2\cos(\omega_m t) + \tfrac{1}{2}(m_{AM} + m_{FM})A_0^2\cos(\omega_m t)$$

or

$$\tfrac{1}{2}A_0^2 + m_{AM}A_0^2\cos(\omega_m t).$$

From this result the original AM component can be recovered easily.

Anderson, J. C. (**1984**). "Speech analysis/synthesis based on perception," Ph.D. Thesis, MIT, Cambridge, MA; MIT Lincoln Lab. Tech. Report 707.

Boashash, B. (**1992a**). "Estimating and interpreting the instantaneous frequency of a signal—Part 1: Fundamentals," Proc. IEEE **80**(4), 520–538.

Boashash, B. (**1992b**). "Estimating and interpreting the instantaneous frequency of a signal—Part 2: Algorithms and applications," Proc. IEEE **80**(4), 540–568.

Carden, F., Jedlicka, R., and Henry, R. (**2002**). *Telemetry Systems Engineering* (Artech, Norwood, MA).

Carmona, R., Hwang, W., and Torrésani, B. (**1997**). "Characterization of signals by the ridges of their wavelet transforms," IEEE Trans. Signal Process. **45**(10), 2586–2590.

Carmona, R., Hwang, W., and Torrésani, B. (**1999**). "Multiridge detection and time-frequency reconstruction," IEEE Trans. Signal Process. **47**(2), 480–492.

Cohen, L. (**1995**). *Time-frequency Analysis* (Prentice Hall, Englewood Cliffs, NJ), pp. 44–52 and 113–175.

Delprat, N., Escudié, B., Guillemain, P., Kronland-Martinet, R., Tchamtichian, P., and Torrésani, B. (**1992**). "Asymptotic Wavelet and Gabor analysis: Extraction of Instantaneous Frequencies," IEEE Trans. Inf. Theory **38**(2), 644–664.

Flanagan, J. L. (**1956**). "Automatic extraction of formant frequencies from continuous speech," J. Acoust. Soc. Am. **28**, 110–118.

Flanagan, J. L. (**1980**). "Parametric coding of speech spectra," J. Acoust. Soc. Am. **68**, 412–419.

Fletcher, N. H. (**1990**). "Nonlinear theory of musical wind instruments," Appl. Acoust. **30**, 85–115.

Fletcher, N. H. (**1998**). "Nonlinearity, complexity, and the sounds of musical instruments," J. Acoust. Soc. Am. **103**(5), 3062.

Gabor, D. (**1946**). "Theory of communication," J. IEE (London) **93**, 429–457.

Harrison, R. V. (**1988**). "The physiology of the cochlear nerve," in *Physiology of the Ear*, edited by A. F. Jahn and J. Santos-Sacchi (Raven, New York), pp. 359–384.

Hertich, I., and Ackermann, H. (**1999**). "A vowel synthesizer based on formant sinusoids modulated by fundamental frequency," J. Acoust. Soc. Am. **106**(5), 2988–2990.

Herzel, H., Holzfuss, J., Kowalik, Z., Pompe, B., and Reuter, R. (**1998**). "Detecting bifurcations in voice signals," in *Nonlinear Analysis of Physiological Data*, edited by H. Kantz, J. Kurths, and G. Mayer-Kress (Springer, Berlin), pp. 325–344.

Houtsma, A. J. M. (**1995**). "Pitch perception," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 267–295.

Jankowski, C. R., Jr., Quatieri, T. F., and Reynolds, D. A. (**1996**). "Fine structure features for speaker identification," *IEEE Int. Conf. On Acoustics, Speech, and Signal Processing, ICASSP-96*, Vol. 2, pp. 689–692.

Koenig, R., Dunn, H., and Lacey, L. (**1946**). "The sound spectrograph," J. Acoust. Soc. Am. **18**, 19–49.

Kootsookos, P. J., Lovell, B. C., and Boashash, B. (**1992**). ''A unified approach to the STFT, TFDs, and instantaneous frequency,'' IEEE Trans. Signal Process. **40**, 1971–1982.
Kumaresan, R., and Rao, A. (**1999**). ''Model-based approach to envelope and positive instantaneous frequency estimation of signals with speech applications,'' J. Acoust. Soc. Am. **105**, 1912–1924.
Ladefoged, P. (**2003**). *Phonetic Data Analysis: An introduction to phonetic fieldwork and instrumental techniques* (Blackwell Publishers, Oxford, UK), Chap. 5, pp. 113–115.
Loughlin, P. J. (**1999**). ''Spectrographic measurement of instantaneous frequency and the time-dependent weighted average instantaneous frequency,'' J. Acoust. Soc. Am. **105**, 264–274.
Maganza, C., Caussé, R., and Laloë, F. (**1986**). ''Bifurcation, period doublings and chaos in clarinet-like systems,'' Europhys. Lett. **1**, 295–302.
Maragos, P., Dimakis, A. G., and Kokinos, I. (**2002**). ''Some advances in nonlinear speech modeling using modulations, fractals, and chaos,'' *14th Int. Conf. on Digital Signal Processing, DSP 2002*, Vol. 1, pp. 325–332.
Maragos, P., Kaiser, J. F., and Quatieri, T. F. (**1993**). ''Energy separation in signal modulations with application to speech analysis,'' IEEE Trans. Signal Process. **41**, 3024–3051.
Nelson, D. J. (**2001**). ''Cross-spectral methods for processing speech,'' J. Acoust. Soc. Am. **110**, 2575–2592.
Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (**1999**). *Discrete-time Signal Processing* (Prentice Hall, Upper Saddle River, NJ), pp. 471 and 714–722.
O'Shaughnessy, D. (**2000**). *Speech Communications: Human and Machine* (IEEE, New York), pp. 191–215.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Pitton, J. W., Wang, K., and Juang, B. H. (**1996**). ''Time-frequency analysis and auditory modeling for automatic recognition of speech,'' Proc. IEEE **84**(9), 1199–1215.
Potamianos, A., and Maragos, P. (**1994**). ''A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation,'' Signal Process. **37**, 95–120.
Potamianos, A., and Maragos, P. (**1999**). ''Speech analysis and synthesis using an AM-FM modulation model,'' Speech Commun. **28**, 195–209.
Potter, R. K., Kopp, G. A., and Green, H. C. (**1947**). *Visible Speech* (Van Nostrand, New York).
Qian, S., and Chen, D. (**1999**). ''Understanding the nature of signals whose power spectra change with time: Joint analysis,'' IEEE Signal Process. Mag. **16**, 52–67.
Quatieri, T. F. (**2002**). *Discrete-time Speech Signal Processing: Principles and Practice* (Prentice Hall, Upper Saddle River, NJ), pp. 175–308, 388–397, 408–412, and 541–594.
Quatieri, T. F., Hanna, T. E., and O'Leary, G. C. (**1997**). ''AM-FM separation using auditory-motivated filters,'' IEEE Trans. Speech Audio Process. **5**(5), 465–480.
Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ), pp. 38–106 and 250–344.
Rao, A., and Kumaresan, R. (**2000**). ''On decomposing speech into modulated components,'' IEEE Trans. Speech Audio Process. **8**(3), 240–254.
Rosen, S., and Howell, P. (**1991**). *Signals and Systems for Speech and Hearing* (Academic, London), pp. 200–235.
Shamma, S. A. (**1985**). ''Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve,'' J. Acoust. Soc. Am. **78**(5), 1622–1632.
Smits, R. (**1994**). ''Accuracy of quasi-stationary analysis of highly dynamic speech signals,'' J. Acoust. Soc. Am. **96**, 3401–3415.
Sorensen, H. V., and Burrus, C. S. (**1988**). ''Efficient computation of the short-time fast Fourier transform,'' in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP-88*, Vol. 3, pp. 1894–1897.
Teager, H. M., and Teager, S. M. (**1989**). ''Evidence for nonlinear sound production mechanisms in the vocal tract,'' in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal, NATO Adv. Study Inst. Series D, Vol. 55.
Thomas, T. J. (**1986**). ''A finite element model of fluid flow in the vocal tract,'' Comput. Speech Lang. **1**, 131–151.
Tolba, H., and O'Shaughnessy, D. (**1998**). ''Automatic speech recognition based on cepstral coefficients and a mel-based discrete energy operator,'' *IEEE Int. Conf. On Acoustics, Speech, and Signal Processing, ICASSP-98*, Vol. 2, pp. 973–976.
Yang, X., Wang, K., and Shamma, S. A. (**1992**). ''Auditory representations of acoustic signals,'' IEEE Trans. Inf. Theory **38**(2), 824–839.

Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)

Walter M. X. Zimmer[a)]
NATO Undersea Research Centre, V.le San Bartolomeo 400, 19138 La Spezia, Italy

Mark P. Johnson, Peter T. Madsen, and Peter L. Tyack
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 2 December 2004; revised 19 March 2005; accepted 24 March 2005)

Strandings of beaked whales of the genera *Ziphius* and *Mesoplodon* have been reported to occur in conjunction with naval sonar use. Detection of the sounds from these elusive whales could reduce the risk of exposure, but descriptions of their vocalizations are at best incomplete. This paper reports quantitative characteristics of clicks from deep-diving Cuvier's beaked whales (*Ziphius cavirostris*) using a unique data set. Two whales in the Ligurian Sea were simultaneously tagged with sound and orientation recording tags, and the dive tracks were reconstructed allowing for derivation of the range and relative aspect between the clicking whales. At depth, the whales produced trains of regular echolocation clicks with mean interclick intervals of 0.43 s (± 0.09) and 0.40 s (± 0.07). The clicks are frequency modulated pulses with durations of ~ 200 μs and center frequencies around 42 kHz, -10 dB bandwidths of 22 kHz, and $Q_{3\,\mathrm{dB}}$ of 4. The sound beam is narrow with an estimated directionality index of more than 25 dB, source levels up to 214 $\mathrm{dB_{pp}}$ *re*: 1 μPa at 1 m, and energy flux density of 164 dB *re*: 1 $\mu\mathrm{Pa}^2$ s. As the spectral and temporal properties are different from those of nonziphiid odontocetes the potential for passive detection is enhanced. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1910225]

PACS numbers: 43.80.Ka [WA] Pages: 3919–3927

I. INTRODUCTION

Since the basic characterization of echolocation clicks of bottlenose dolphins by Au *et al.* (1974), it has been known that toothed whales are capable of producing highly directional ultrasonic clicks with source levels of more than 220 $\mathrm{dB_{pp}}$ *re*: 1 μPa at 1 m. During 30 years of studies on captive animals (Au, 1993), and increasingly during the last decade on free-ranging animals (Møhl *et al.*, 1990; Au and Herzing, 2003; Schotten *et al.*, 2003; Au *et al.*, 2004; Madsen *et al.*, 2004), it has become apparent that toothed whale species produce a diverse range of biosonar signals (Au, 1997). Overall, toothed whale sonar signals can be divided into the low-output, monochromatic, high frequency pulses of *Phocoena* and *Cephalorhynchus* and the shorter more broadband and higher source level clicks of most dolphin species (Au, 1997). A third group, sperm whales, produces multi-pulse sonar clicks, dominated by a single highly directional pulse centered around 15 kHz with source levels of more than 230 $\mathrm{dB_{rms}}$ *re*: 1 μPa at 1 m (Møhl *et al.*, 2003; Zimmer *et al.*, 2005).

Information on the characteristics of sonar signals from a major group of some 20 species of deep-diving toothed whales, the beaked whales (*Ziphiidae*) (Rice, 1998), is sparse. There are a few recordings of sounds from stranded whales (Caldwell and Caldwell, 1971; Lynn and Reiss, 1992; Marten, 2000) that may not be representative of the sounds of healthy, free ranging whales. Data from free-ranging beaked whales in offshore habitats is limited to a few species, including *Hyperoodon* (Hooker and Whitehead, 2002) and *Berardius* (Dawson *et al.*, 1998).

This lack of information is especially serious for Cuvier's beaked whale (*Ziphius cavirostris*), as this species appears to be highly sensitive to anthropogenic noise as demonstrated by several mass strandings during and after military sonar exercises (Simmonds and Lopez-Jurado, 1991; Frantzis, 1998; Jepson *et al.*, 2003). Quantitative data on how *Ziphius* produce and use sound may help in understanding the sensitivity of this species to anthropogenic sound and will be crucial for the development of passive identification and monitoring techniques, to minimize the impact of human activity.

Frantzis *et al.* (2002) analyzed about 5 h of recordings acquired off SW Crete in proximity to *Ziphius*. They reported the presence of click trains with average click duration of 1.08 ms and with click energy concentrated in a narrow peak between 13 and 17 kHz. The average interclick interval was 0.44 s, and clicking was frequently interrupted by short pauses. Frantzis *et al.* (2002) did not detect click sequences with elevated click rates and speculated that the whales may use visual cues in the final stages of prey capture.

Recently, Johnson *et al.* (2004) reported data from the successful tagging of four beaked whales [two Blainville's beaked whales (*Mesoplodon densirostris*) and two *Ziphius*] demonstrating echoes from prey recorded after the whales produced ultrasonic clicks for echolocation. Like other odontocetes and bats, *Ziphius* produce high repetition click trains (so-called buzzes) during the final stages of prey capture (Johnson *et al.*, 2004). The intervals between regular clicks for *Ziphius* were close to 0.4 s and the clicks were ~ 175 μs

a)Electronic mail: walter@nurc.nato.int

long with a relatively flat spectrum from 30 kHz up to the 48 kHz Nyquist cutoff frequency of the acoustic recording system in use. No information could be given on the high-frequency limit of click energy or the radiation pattern of the clicks. Such data are necessary to study the biosonar performance of this little known species and to provide a quantitative basis for evaluating passive acoustic detection and monitoring.

In this paper we report estimates of the source characteristics of *Ziphius* echolocation clicks using the novel approach of two tagged whales recording each other. We provide the first estimates of the source level, directivity index, and spectral properties of *Ziphius* clicks, which are shown to differ significantly from those produced by non-ziphiid toothed whales described so far, suggesting a strong potential for passive acoustic monitoring.

II. METHODS

The results in this paper are based on three data sets from sea trials performed in the Ligurian Sea by the NATO Undersea Research Center (NURC) and the Woods Hole Oceanographic Institution (WHOI) as part of their combined effort to broaden the knowledge on beaked whales to mitigate detrimental effects of human activity.

(1) July 2002 NURC: recordings from a single hydrophone with 384 kHz sampling rate.
(2) June 2004 WHOI: recordings from digital recording tags (DTAG) on two whales in a group of 5 (zc04_161a, zc04_161b). The sampling rate was 96 kHz and the tags remained attached for 9 and 19 h, respectively.
(3) June 2004 WHOI: stereo DTAG recording from a single whale in a group of 4 *Ziphius* (zc04_179a). The sampling rate was 192 kHz and the tag remained attached for 4 h.

The NURC July 2002 sea trial (*Sirena*02) took place in an area of underwater canyons known to be inhabited by *Ziphius*. Following standard oceanographic and biological sampling, and whenever *Ziphius* were sighted, a hydrophone was deployed to about 80 m for 1 h of acoustic recording. The June 2004 WHOI experiment was dedicated to tagging and took place in the same area as the NURC July 2002 sea trial. After sightings of *Ziphius*, the tagging team slowly approached the whales in a small inflatable boat. A lightweight 5 m carbon fiber pole was used to deliver the tag. The tag was attached to the whale with a set of four small silicone suction cups that were vented after a programmable time to release the tag from the whale. The tag was then recovered with the aid of a built-in VHF transmitter (Johnson and Tyack, 2003).

A. Instrumentation

Two acoustic receivers were used to collect the data presented in this paper. The July 2002 NURC sea trial used a single broadband hydrophone (-3 dB bandwidth from 1 to 160 kHz) with a NURC-designed low noise pre-amplifier (spectral electronic noise level 15 dB *re*: 1 μPa^2/Hz at 50 kHz). The hydrophone signal was recorded by a Sony SIR1000 recorder sampling at 384 kHz. The overall clipping level was 140 dB_{peak} *re*: 1 μPa and the data were digitized with 16 bit resolution.

The DTAG, a miniature sound and orientation recording tag developed at WHOI, was used during the June 2004 WHOI sea trial. Mono or stereo acoustic data were sampled at 96 or 192 kHz using a 16-bit resolution sigma-delta analog digital converter. The clipping level was set to 181 (96 kHz version) and 171 (192 kHz version) dB_{peak} *re*: 1 μPa. Accelerometers, magnetometers, and a pressure sensor were sampled at 50 Hz to measure orientation and depth of the tagged whale. Data were stored digitally in up to 6.6 Gbyte of nonvolatile memory (Johnson and Tyack, 2003).

B. Processing

During deep dives, the dominant sounds in the tag recordings were clicks from the tagged whale and nearby conspecifics. Clicks from the tagged whale and conspecifics can be distinguished spectrally: due to the physical attachment of the tag to the body of the whale there is relatively high coupling of sound energy below 20 kHz from that whale to the tag. This spectral energy is virtually absent in the clicks from conspecifics. Based on these observations, analysis of the recordings proceeded in several steps.

Clicks from each tagged whale were automatically detected and verified by visual inspection of temporal and spectral characteristics, to establish the time of click emission and the interclick interval (ICI) statistics. The position of the tags on the dorsal surface of the whale, behind the sound source, leads to a poor estimate of the spectrum of the clicks in the forward direction. For this reason, the tag recordings were examined for clicks from other whales with ICI and spectral characteristics consistent with far-field *Ziphius* clicks (Johnson *et al.*, 2004). These were checked individually and the process was repeated for the NURC recording. The result was a set of far-field *Ziphius* click wave forms with unknown source aspect. Many of these clicks occurred in short sequences of increasing and then decreasing amplitude, presumably due to movement of the clicking whale with respect to the receiver as the whale scanned with a directional sonar beam. Assuming that *Ziphius*, like all other odontoceti investigated, concentrate sound energy into a forward-directed beam, the highest amplitude clicks were considered to best represent the click wave form within the beam (Møhl *et al.*, 2000; Johnson *et al.*, 2004). These clicks were reserved for spectral analysis.

In addition to the power spectral density, signals were parametrized in terms of energy flux density and mean or center frequency. Energy flux density is estimated by the formula

$$E = \frac{1}{\rho c} \int_{T_1}^{T_2} |s(t)|^2 dt, \tag{1}$$

where $s(t)$ is the pressure time-series of the click and the integration boundaries T_1 and T_2 are determined such that E covers 97% of the measured signal energy. The term ρc is the characteristic impedance of the propagation medium, which for water is $\rho c \approx 1.5 \times 10^6$ rayl (Lurton, 2002). The

signal duration is then determined by the time difference between the integration boundaries:

$$\tau = T_2 - T_1. \tag{2}$$

A key parameter used to describe the spectral characteristic is the mean or center frequency of the spectrum, which was estimated by the formula (Au, 1993):

$$f_0 = \frac{\int_{-\infty}^{\infty} f|S(f)|^2 df}{\int_{-\infty}^{\infty} |S(f)|^2 df}, \tag{3}$$

where $S(f)$ is the Fourier transform of signal time series.

Two further processing steps were required to estimate the source level (SL) and directivity index (DI) from the simultaneous tag recordings zc04_161a and b. These tags were applied to two whales swimming within the same group and the tag recordings overlapped for 8 h, spanning four deep dives. Having detected the clicks from each of the tagged whales in the corresponding tag, we examined the tag recordings for clicks emanating from the other tagged whale, i.e., recording B (zc04_161b) was examined for clicks from whale A (zc04_161a) and *vice versa*. This was achieved by cross-correlating the times of clicks made by one tagged whale with the reception times of conspecific clicks in the other tag recording. The time delays were noted in each direction (i.e., from whale A to whale B and from B to A) and the range and clock offset between the tags were estimated.

In order to relate the received level of clicks to the relative position of the clicking whale, the underwater track of each whale was reconstructed. This was achieved by combining the orientations recorded by the tags with swim-speed estimates, and the predicted ranges between the two whales. The orientation of each whale was obtained from the accelerometers and magnetometers in the DTAG (Johnson and Tyack, 2003; Zimmer *et al.*, 2003), corrected for the position of the tag on the body. The orientation of the tag on the whale was estimated from the data by assuming that while the whale is swimming on the surface, the mean pitch and roll are zero and the variance of the roll is minimal. The procedure further assumes that the mean swimming direction is parallel to the body axis. The swim-speed estimator comprised a Kalman filter (Jazwinski, 1970) to fit the pitch angle to the depth profile recorded by the tag. Combining the estimates, the track of each whale, up to integration constants, was obtained by integrating the (x, y, z) components of orientation multiplied by the swim speed. The acoustic ranges were finally used to improve the tracks and to obtain the integration constants.

Knowing the separation and relative orientation of the two whales, it was possible to estimate the elevation and azimuth of each whale with respect to the receiving whale for all clicks. These data were used to calculate the apparent source level (ASL *sensu* Møhl *et al.*, 2000) as a function of off-axis angle, i.e., the angle between the caudal-rostral axis of the clicking whale and the vector joining the two whales. The transmission loss due to sound propagation between the whales was estimated using spherical spreading and an absorption loss of 10 dB/km for a frequency of 40 kHz (Lurton, 2002). Source level (SL) and DI were then obtained by analyzing the shape of the ASL function. In reality, the off-axis angle is relative to the body axis, while the description of the acoustic beam requires knowledge of the acoustic axis (Au *et al.*, 1986). The analysis of the measured ASL function must therefore allow for differences between acoustic and body axes.

As in other cases (Au, 1993; Møhl *et al.*, 2003; Rasmussen *et al.*, 2004; Zimmer *et al.*, 2005), the acoustic beam of odontocetes may be described by a flat circular piston model:

$$P(x) = P_0 \frac{2J_1(x)}{x}, \tag{4}$$

with

$$x = ka \sin\vartheta = 2\pi \frac{a \sin(\vartheta)}{c} f,$$

where P_0 = source level, a = piston radius, c = speed of sound, ϑ = off-axis angle, f = frequency, J_1 = Bessel function of the first kind.

This model provides a convenient way to parametrize the ASL function using a minimum set of parameters, namely, source level, effective piston radius, frequency, and off-axis angle. The broadband beam pattern, $B(\vartheta)$, can be approximated by integrating $P(x)$ with respect to frequency. A spectrum based weighting function $W(f)$ is used to account for the variation in source level with frequency,

$$B(\vartheta) = \frac{\int_{-\infty}^{\infty} P^2(\vartheta, f) W^2(f) df}{\int_{-\infty}^{\infty} W^2(f) df}. \tag{5}$$

As will be shown, the power spectrum of *Ziphius* clicks can be approximated by a Gaussian function, which will be used for the weighting function:

$$W(f) = \exp\left\{ -\frac{1}{2}\left(\frac{f - f_0}{b} \right)^2 \right\}, \tag{6}$$

where f_0 and b are the center frequency and rms bandwidth, respectively, and obtained by the least-mean-square fit of $W(f)$ to the measured power spectrum.

The directivity index is finally obtained by numeric integration of the modeled broadband beam pattern (Lurton, 2002):

$$\mathrm{DI} = 10 \log\left(\frac{\mathrm{B}(0) \int_0^{\pi} \sin\vartheta \, d\vartheta}{\int_0^{\pi} \mathrm{B}(\vartheta) \sin\vartheta \, d\vartheta} \right). \tag{7}$$

The directivity index may be related to the beam width by combining the expression for the -3 dB beam width $\Theta = 2\vartheta_{-3\,\mathrm{dB}} \approx 185°/ka$ with the expression for the directivity index $\mathrm{DI} \approx 20 \log(ka)$, which are valid for a circular piston and $ka \gg 1$ (Lurton, 2002), to obtain the following relation:

$$\Theta \approx 185° \times 10^{-\mathrm{DI}/20}. \tag{8}$$

III. RESULTS

Echolocation clicks from tagged whales and conspecifics were reliably detected in all of the tag recordings. In the double tag data set (zc04_161a and b), clicks made by one of the tagged whales were frequently heard in the other tag, and *vice versa*, providing sufficient estimates of range to perform

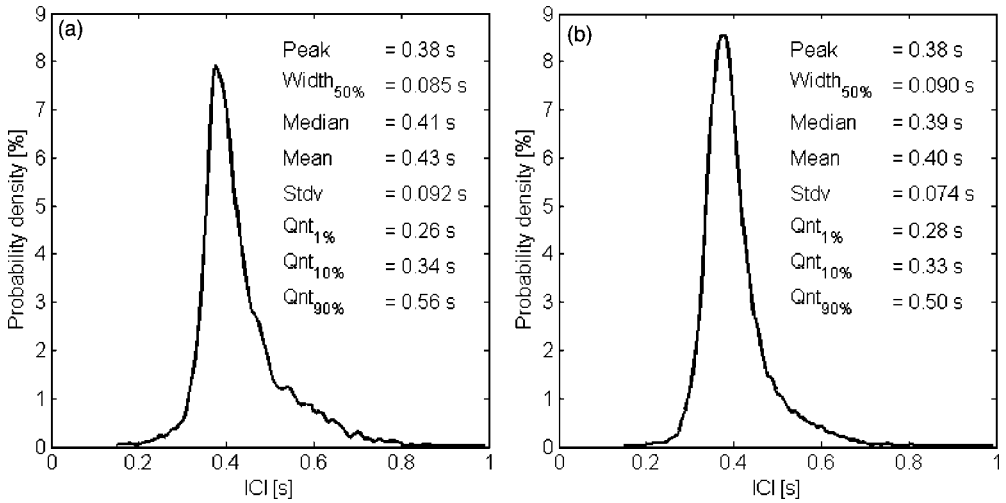


FIG. 1. Probability density function of the inter-click interval (ICI) of both whales. A total of 10 736 and 11 117 of regular clicks (ICI$>$0.15 s and ICI $<$1 s) were used for whales A and B, respectively. The bin width for the density function was 0.005 s, and the bins with the largest number of ICIs were 0.38 s for whales A and B. The density function does not include the ICI of terminal click buzzes with ICI$<$0.15 s and over extended pauses with resulting ICI $>$1 s.

track reconstruction. Far-field *Ziphius* clicks were also detected in the recording made with the NURC hydrophone.

A. Temporal and spectral characteristics

Interclick intervals (ICIs) were measured for two whales called whale A and whale B. Figure 1 shows for both whales the probability density function of the ICI over three dives. The estimates do not include the ICI of buzz sounds with ICI$<$0.15 s or the occasional pauses in clicking with ICI values$>$1 s. Buzzes are characterized by a significant decrease in source level and consequently the same buzz clicks were not reliably detected on both whales. For whale A, 10 736 clicks were analyzed for ICI, yielding a mean ICI of 0.43 s with a standard deviation of 0.092 s and a median of 0.41 s. For whale B, 11 117 clicks were analyzed for ICI, yielding a mean ICI of 0.40 s with a standard deviation of 0.074 s and a median of 0.39 s. Both histograms are asymmetric with a sharp limit on the lower side and softer decay at higher ICI.

Figure 2 shows the time series, power spectrum, and time-frequency spectrogram (short-time Fourier transform with window size of 32 samples, and 31 samples overlap) of a representative far-field *Ziphius* click from the zc04_179a recording with 192 kHz sampling rate. The effective duration of the click is based on the 97% energy criteria and is estimated as $\tau = 203\ \mu$s [Eq. (2)]. The spectrogram indicates that the signal is frequency modulated with a half power frequency range starting from 35 kHz and ending around 45 kHz and with a center frequency of 42 kHz. Similar spectral and temporal characteristics were observed in all far-field clicks, notwithstanding the power spectra fluctuations (especially at higher frequencies) due to variations in the aspect of the clicking whales.

Far-field click recordings at sampling rates of 192 and 384 kHz were examined to assess the adequacy of the 96 kHz sampling rate used in the double tag recording. Although this reduced bandwidth does not fully cover the spectrum of the clicks, it is important to determine whether it is sufficient to sample most of the click energy and therefore would yield reliable estimates of ASL and DI. Far field clicks from different whales with different recording bandwidths are compared in Fig. 3. All three clicks were selected from sequences that probably describe a scan of a *Ziphius* that passes the recording hydrophone; that is, within 10–20 clicks

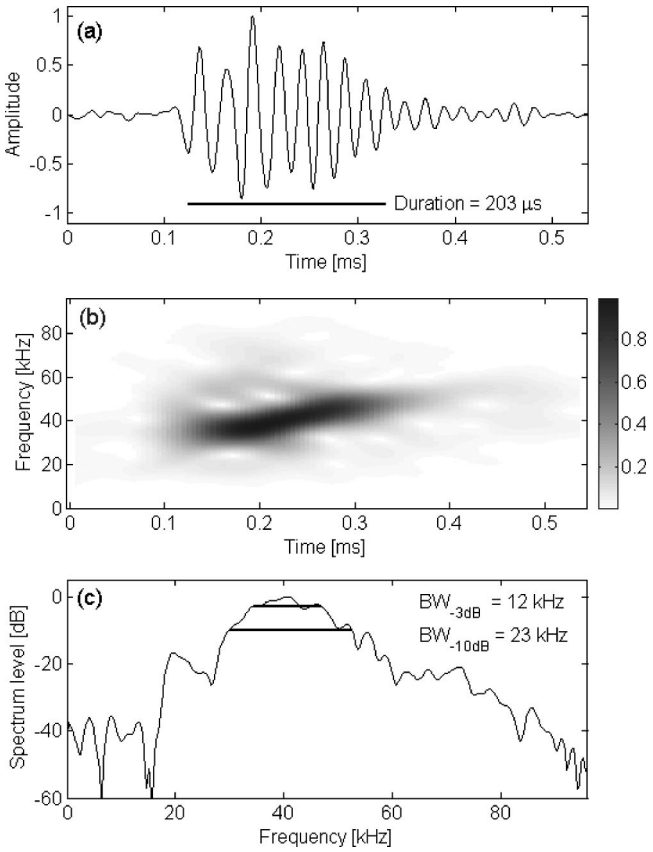


FIG. 2. Sample click of *Ziphius* click as received by the tag of a conspecific whale. The top panel shows the time series, the middle panel shows the spectrogram (linear spectral scale, Hann window for FFT of 32 samples, and overlap of 31 samples) and bottom panel shows the relative power spectrum.

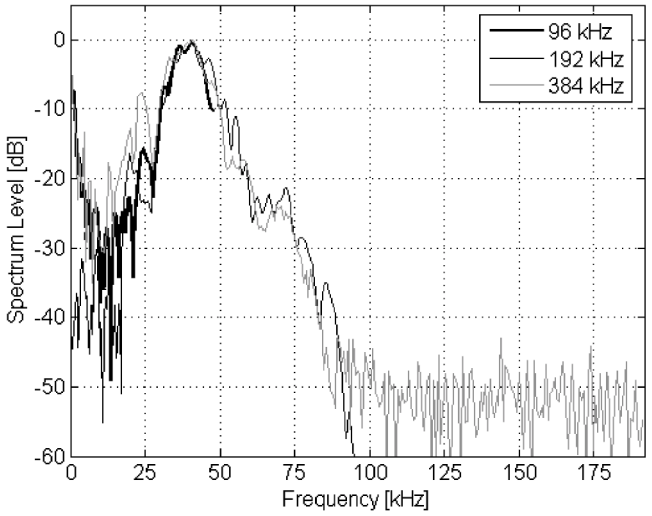


FIG. 3. Spectra of three different *Ziphius* clicks with three different sampling frequencies. The spectral levels are relative to the peak level at 40 kHz.

the received signal level varied from weak to strong and to weak again. It was assumed that the click with the highest signal level was closest to the acoustic axis (*sensu* Møhl *et al.*, 2003) and therefore suited for comparison. Between 26 and 60 kHz, the spectrum sampled at 192 kHz may be fitted by a Gaussian function [Eq. (6)] with a center frequency $f_0 = 42.1$ kHz and rms bandwidth $b = 7.9$ kHz. Using this model, the 96 kHz recording with a -3 dB bandwidth of 47.5 kHz, samples $\sim 80\%$ of the energy in the click and is therefore suitable for SL and DI analysis.

B. Source level and directivity

The relative orientation and range between source and receiver was estimated, prior to source level and directivity estimation.

1. Reconstruction

The top panel of Fig. 4 shows the reconstructed horizontal track for a simultaneous dive of both whales. It can be seen that the swimming behavior is coordinated: the two whales dived close together, moved on similar tracks while underwater, then approached each other and surfaced with a nearly parallel track. The phase selected for estimation of source level and directivity is indicated with a bolder line on both tracks. Clicks from whale A were detected during this phase on the tag attached to whale B. The bottom panel shows the range estimate from the reconstruction for this selected phase. The independent acoustic range estimates based on the travel time of clicks are overlaid and marked with dots for clicks from whale A to whale B and with triangles for clicks from whale B to whale A. The ranges for the analyzed clicks vary between 400 and 100 m.

2. Click beam pattern

Figure 5 shows the ASL of all clicks produced by whale A and received by whale B during the selected phase. The received level at the receiver is corrected for spherical spreading and attenuation loss. The azimuth and elevation

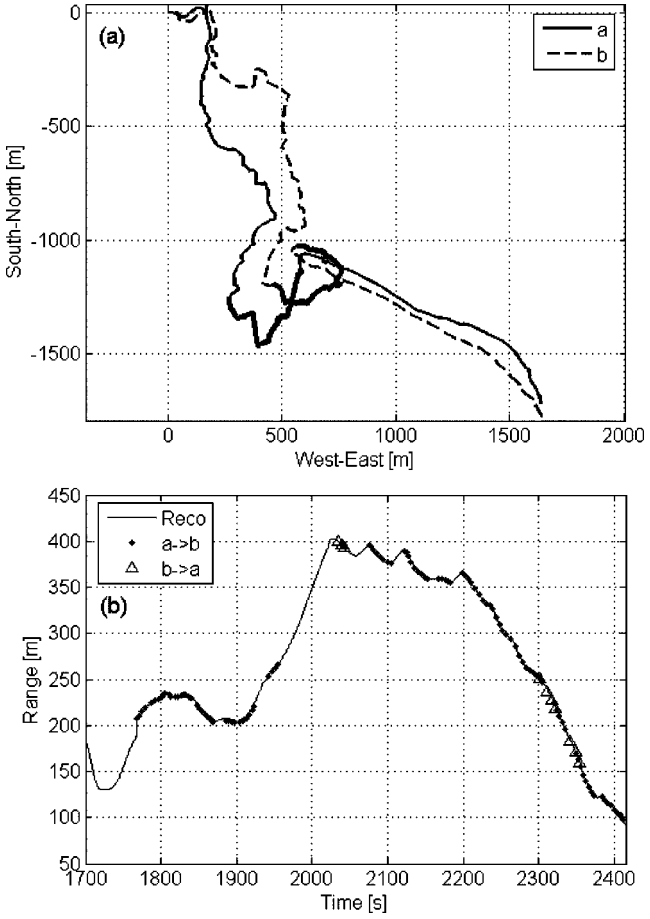


FIG. 4. Reconstruction of the tracks of two *Ziphius* carrying tags at the same time. Top panel: Plan view of horizontal components of tracks of whale A (solid line) and whale B (dashed line). The bold portions of the tracks mark the period when whale A approached whale B and the clicks of one tagged whale were also audible on the tag of the other whales. Bottom panel: Range between the two tagged whales. Each marker represents an acoustic range estimate; dots describe ranges for clicks emitted by whale A and received by whale B, and triangles correspond to ranges for clicks from whale B that were received by whale A.

angles of the clicking whale are plotted from the transmitter's point of view (whale A). Off-axis angles are drawn as circles around the forward orientation, i.e., where azimuth and elevation are zero. The scatter plot shows that the data available are not sufficient to describe a complete three-dimensional beam pattern of the transmitted sound energy. Apart from some traces with off-axis angles $>60°$, most clicks are received near the center with off-axis angles $<40°$. There is considerable variation in the directions of the strongest clicks (>190 dB *re*: 1 μPa) indicating that there is no unique and preferred direction for the click emission relative to the body axis. On the other hand, these strong clicks are isolated, which may reflect variations in the source level produced by the tagged whale as well as steering of a narrow beam by head movements altering the relation between the body axis measured by the tag and the acoustic axis of the whale.

Figure 6 shows the apparent source level (ASL) of a single scan as a function of time (left) and off-axis angle (right). A level variation of 30 dB during this scan makes it well suited to the estimation of source level (SL) and direc-

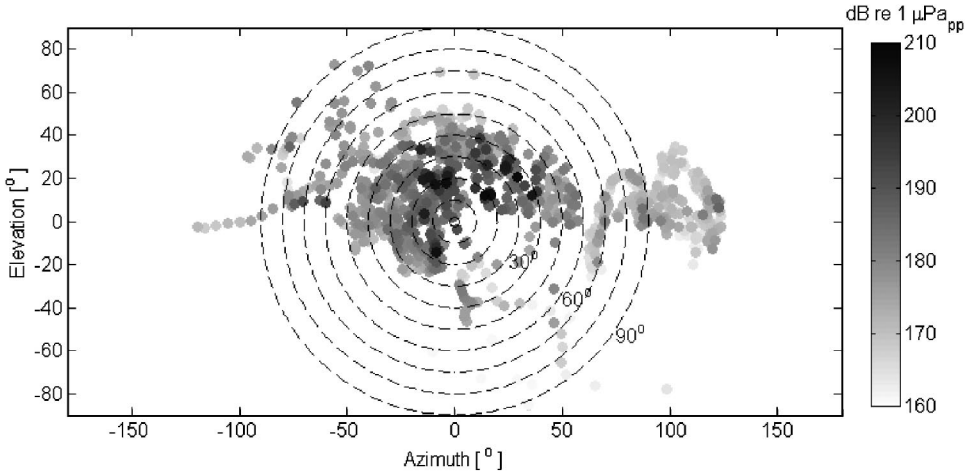


FIG. 5. Two-dimensional scatter plot of apparent source level as function of azimuth and elevation for the 1123 clicks emitted by the tagged whale and received by the companion. The angles are plotted from the transmitter's point of view. The overlaid circles correspond to the off-axis angles between the body axis of the transmitting whale and the position of the receiving whale.

tivity index (DI). Inspection of the click levels at the transmitting whale revealed a variation of less than 3 dB between 2074 and 2078 s, and less than 4 dB between 2078 and 2088 s, suggesting that the source level is stable during this scan (Madsen *et al.*, 2004). The variation in received level at the ''scanned'' whale is thus likely to be the result of a constant output, directional source that moves past the receiver, the maximum ASL should be a reasonable proxy for source level.

The measurements in Fig. 6 are marked by a triangle during the period of increasing received level and an asterisk during the period of decreasing received level to emphasize the asymmetry of the scan. While the decrease of the ASL (''*'') coincides with an increase of the off-axis angle [marked by an open circle (○)], the sharp increase ASL before its maximum cannot be explained by the off-axis angle that remains nearly constant. This is made clearer in the right panel where a broadband piston model, driven by the measured off-axis angle is superimposed. The least mean square fit of the broadband piston model to the decreasing ASL values resulted in a piston diameter $d=40$ cm, which is equivalent to a broadband directivity index DI$=30$ dB when radiating a *Ziphius* click [Eq. (7)], corresponding in turn to a -3 dB beam width of $\Theta=6°$ [Eq. (8)]. The modeled piston beam (marked with open circles) has its maximum at an off-axis angle of 15° indicating that, for this scan, the body axis and the acoustic axis are not aligned, which may relate to head movements and a possible offset between body and acoustic axis as seen in dolphins (Au, 1993). The temporal and spectral values of the *Ziphius* click at the maximum of the scan are compiled in Table I, which also provides a comparison with three other echolocating toothed whales.

3. Off-axis click distribution

All ASL values are given again in Fig. 7, which shows in gray the ASL as a function of off-axis angle. Superimposed are the data from the single scan of Fig. 6 [marked ''*''], and the modeled broadband piston beam (solid line) that was fitted to the selected scan. The piston beam not only fits the selected scan but also seems to be a fair approximation of most clicks below an ASL of 190 dB.

IV. DISCUSSION

A. Temporal and spectral characteristics

The echolocation clicks of *Ziphius* occur regularly with an ICI averaging about 0.4 s. Short pauses are frequent and result in an asymmetric distribution of the ICI as shown in Fig. 1. If the two whales tagged are representative of the population, the sharp peak of the distribution around 0.4 s suggests that lengthy regular click trains (>10 s) with ICI less than 0.26 s are with 99% probability not being made by

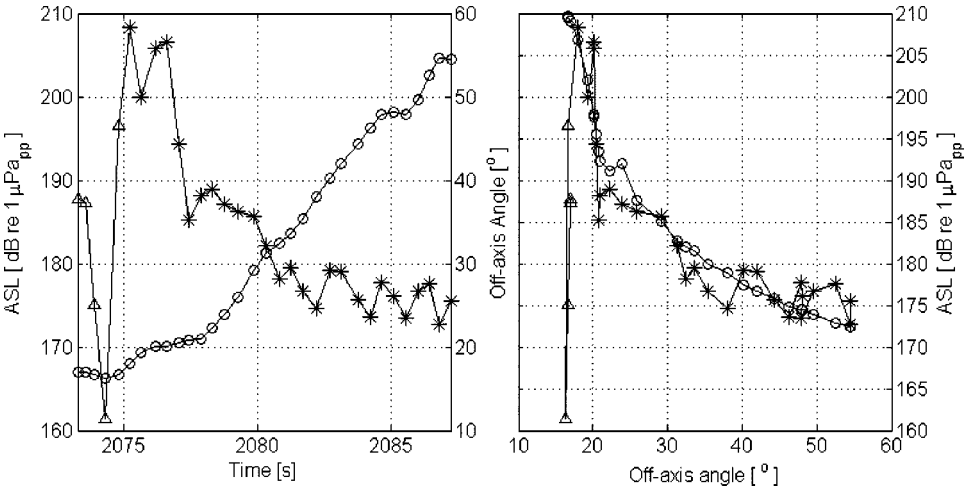


FIG. 6. Single scan of whale A as recorded by whale B. The left panel shows as function of time (a) the ASL (triangles and asterisks) and the off-axis angle (open circles). The right panel shows as function of off-axis angle (a) the ALS (triangles and asterisks) and (b) a prediction (open circles) of the ASL by a piston model.

TABLE I. Comparison of salient parameters for the description of echolocation signals of four toothed whales, harbor porpoise (*Phocoena phocoena*), bottlenose dolphin (*Tursiops truncatus*), Cuvier's beaked whale (*Ziphius cavirostris*), and sperm whale (*Physeter macrocephalus*).

	Phocoena (Au *et al.*, 1999)	*Tursiops* (Au, 1993)	*Ziphius* (this paper)	*Physeter* (Møhl *et al.*, 2003; Madsen *et al.*, 2002)
SL_{pp} [dB re: 1 μPa at 1 m]	170	228	**214**	240
Energy flux density [dB re: 1 μPa^2 s]	130	167	**164**	195
DI [dB]	22	26	**>25**	27
Duration [μs]	100	25	**200**	120
F_{peak} [kHz]	130	120	**40**	15
F_0 [kHz]	135	100	**42**	20
−10 dB BW [kHz]	20	100	**23**	10
−3 dB BW [kHz]	10	30	**12**	5
$Q = Fp/BW_{-3\,dB}$	13	2–3	**4**	2–3

Ziphius, if one ignores buzzes that have ICI≪0.15 s (Johnson *et al.*, 2004) and are difficult to detect when recording in the far field.

The spectra of *Ziphius* clicks peak at 40 kHz with −30 dB points in energy ranging from 15 to 80 kHz, and −10 and −3 dB bandwidths of around 23 and 12 kHz, respectively. The *Q* of the clicks (i.e., the center frequency divided by the −3 dB bandwidth) is about 4, and is closer to that of dolphin and sperm whale clicks ($Q = 2-3$) than it is to species such as porpoise that produce monochromatic signals with *Q*'s of more than 10 (Table I).

Representative spectra for *Ziphius* clicks, recorded by three different instruments between 2002 and 2004, are shown in Fig. 3. Although these clicks are almost certainly from different whales, all three spectra have a spectral peak at 40 kHz and a spectral notch at about 26 kHz. Spectral differences above the peak frequency may indicate that the measurements were not all made precisely on the acoustic axis of the clicking whale and so may include some off-axis distortion (Au, 1993). The features just below the 26 kHz notch also differ for the three different plots. This may also be due to off-axis distortion, but could be a characteristic of individual whales.

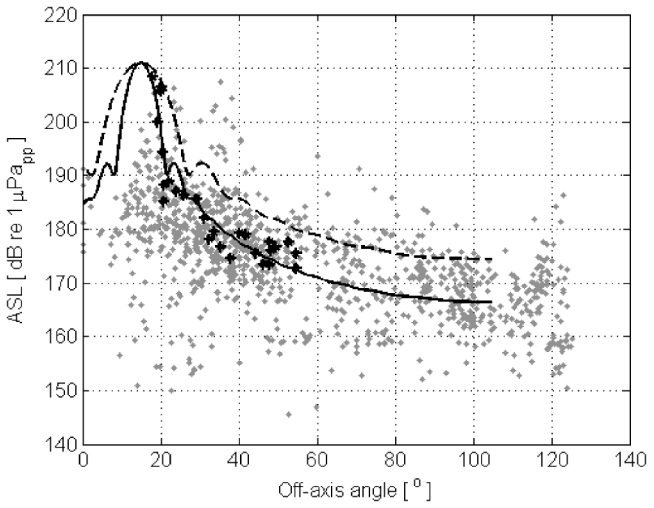


FIG. 7. The apparent source level (ASL) of 1123 clicks is plotted in gray against off-axis angle. In black are the clicks from single scan (the asterisks from Fig. 6). The solid line is the broadband beam pattern of the modeled piston with piston diameter 0.40 m (equivalent broadband DI = 30 dB), and off-axis angle of acoustic axis of 15°, the dashed line describes a broadband beam pattern for a DI = 25 dB.

The click duration of around 200 μs warrants some discussion as this is considerably longer than the duration of clicks of any non-ziphiid toothed whale. Most delphinids produce clicks with durations between 20 (Rasmussen *et al.*, 2004) and 100 μs (Au *et al.*, 2004) so the echolocation clicks of *Ziphius* have a duration that is at least twice as long. The sperm whale produces multi-pulsed clicks with an overall longer duration, but the dominant P1 pulse has durations around 120 μs (Møhl *et al.*, 2003). Comparing the duration of the individual pulses that make up the sperm whale regular click, it is evident that the clicks of *Ziphius* are considerably longer (Table I).

Thus, the clicks of *Ziphius* differ from clicks recorded from delphinids and sperm whales in combining a long duration with center frequencies around 40 kHz. Killer whales (Au *et al.*, 2004) and narwhals (Møhl *et al.*, 1990) have similar center frequencies, but much shorter durations and lower *Q*'s. The long click duration is a prerequisite for the frequency modulated (FM) sweep seen in *Ziphius* that sweeps from a frequency of about 35–45 kHz.

When comparing the click properties of *Ziphius* to those reported for the Northern Bottlenose whale (*Hyperoodon ampulatus*) by Hooker and Whitehead (2002) using a band limited (<40 kHz) single hydrophone system, it appears that the *Ziphius* clicks are of higher frequency and of longer duration than clicks from the larger *Hyperoodon*. It is not clear if such apparent differences relate to differences in sound production, biosonar performance, size of the whale, orientation of the whales with respect to the hydrophones or simply differences in recording equipment and setup.

B. Source level and directivity

Analysis of the broadband click spectrum shows that the sampling rate of the double tag data, 96 kHz, is suitable for estimating source level and directivity index. To obtain source level (SL) and directivity index (DI), the angle of the receiver with respect to the acoustic axis (the "off-axis

angle'') must be estimated for each click. However, the off-axis angle, as derived from the tag data, is really the angle between the joining line to the receiver and the direction of motion of the clicking whale. While it cannot be excluded that the whale swam with a pitch offset, it has been assumed that the mean body axis is parallel to the mean direction of motion, so that the body axis is on average parallel to the mean swim direction. However, it seems most likely that the whale, while clicking, is moving the head and thereby the acoustic axis, separately from the body axis. This notion is supported by visual observations of whales at the surface and by anatomical evidence based on a pivot point at the occipital condyles/atlas-axis and partly unfused cervicals (Allen and Mead, private communication) suggesting that *Ziphius* has ample room for head motion while clicking. Finally, the authors have observed significant head movements in another beaked whale species (*Mesoplodon densirostris*) while clicking, supporting the notion that this is also likely the case in *Ziphius*.

Figure 7 has a striking lack of clicks recorded at off-axis body angles $<10°$ whereas most strong values (ASL >190 dB *re*: 1 μPa) are between 15° and 40°. From Fig. 5 it may be deduced that the measurements do not cover all aspects (e.g., there are no data for azimuth $10°-30°$ and elevation $0°-20°$) and that the strong clicks are scattered around the forward direction. Accordingly, some of the scatter in the data of Figs. 5 and 7 is likely the result of head movements by the clicking whale. Likewise the lack of recorded clicks from off axis body angles $<10°$ may be a consequence of inadequate sampling. The increasing part of the selected scan, marked by triangles in Fig. 6, appears to be independent of the off-axis angle variation due to body motion, suggesting that the increase in level represents a head scan. The sharp increase of the received level during this apparent scan suggests that excursion of the head movement is at least 25° with a scan rate of 25°/s, which is consistent with observations of Frantzis *et al.* (2002). The correlation of ASL with body axis during the decreasing part of the scan (marked ''*'') suggests that body motion dominates this effect.

The highest ASL levels are the best candidates for defining the maximum source level (SL) of this species. The highest measured SL of 214 dB_{pp} *re*: 1 μPa at 1 m is considerably lower than the maximum levels of more than 220 dB_{pp} *re*: 1 μPa at 1 m reported for a range of dolphin species (Au, 1993) and at least 20 dB lower than the SL of sperm whale clicks (Møhl *et al.*, 2003; Zimmer *et al.*, 2005). Although it is possible that *Ziphius* cannot produce higher SLs, it is more likely that the full capabilities of *Ziphius* are underestimated here. The data set of 1123 measurements from a single dive may also represent a biased data set for SL estimation. A potential problem with using tags to measure the SL of clicks from conspecifics is that whales may avoid ensonifying each other with high-powered clicks as has been suggested for spinner dolphins (Brownlee and Norris, 1994). Thus, the maximum of 214 dB_{pp} *re*: 1 μPa at 1 m may simply be the result of the whales reducing the volume when ensonifying each other.

The longer duration of *Ziphius* clicks means that they carry more energy than the clicks of other odontocete species with the same peak-to-peak sound pressure level. The observed energy flux of up to 164 dB *re*: 1 μPa^2 s in *Ziphius* clicks is comparable with the energy flux of clicks of the bottlenose dolphin (*Tursiops truncatus*), even though the measured *Ziphius* source sound pressure level of 214 dB_{pp} *re*: 1 μPa at 1 m is about 15 dB less than the maximum SL of *Tursiops* clicks (Au, 1993).

Au *et al.* (1999) proposed how DI might scale in toothed whales. According to this scaling, a *Ziphius* with head diameter at the blowhole of 60 cm (Allen and Mead, private communication) and a wavelength of 3.3 cm should have a (narrowband) DI of 24.3 dB, and -3 dB beam width of 12.6°. A similar figure is obtained by scaling the results reported by Au *et al.* (1995) for the DI of a false killer whale *Pseudorca crassidens* with a head diameter of 40 cm (measured DI of 22.3 dB at 44.3 kHz). Making the crude assumption that the diameter of the head at the eye scales with the transmitting aperture, we estimate that a *Ziphius* with a head diameter of 60 cm will have a (narrowband) DI of $22.3 + 20\log(60/40\text{ cm}) = 25.8$ dB around 40 kHz, which is about the same as the DI of *Tursiops* clicks radiating from a smaller equivalent aperture, but at a higher frequency. The estimated DI's are about 4–6 dB less than the 30 dB DI derived here using a broadband piston model to fit the measured ASL pattern. It is possible that our DI is an overestimation based on the assumption that the variation in ASL is only due to the angle from the body axis. *Ziphius* appear to move their heads regularly while searching for prey and some head motion cannot be excluded in the selected data segment. Based on the above-noted predictions and the data presented here, it seems reasonable to suggest that *Ziphius* clicks are radiated with a broadband DI of more than 25 dB. Curves for DI = 25 dB and 30 dB are shown in Fig. 7 and appear to bracket the data fairly well.

In conclusion, we have demonstrated that *Ziphius* produce ultrasonic echolocation clicks with center frequencies around 42 kHz and the distinctive form of a FM up-sweep. The derived directivity index of 30 dB is a little higher than predictions based on other toothed whales, but may be a slight overestimation. The maximum SL of 214 dB_{pp} *re*: 1 μPa at 1 m is probably an underestimate. The clicks have a longer duration than clicks from other non-ziphiid toothed whales. The estimated energy flux density of 164 dB *re*: 1 μPa^2 s is comparable to that of clicks from bottlenose dolphins, which have considerably higher peak-to-peak sound pressures. It is thus evident that *Ziphius* produce clicks with temporal and spectral properties that differ from those of clicks produced by most other toothed whales, and that the clicks, on that basis, hold a potential for acoustic classification and passive monitoring.

ACKNOWLEDGMENTS

The authors thank D. Allen and J. Mead from the Smithsonian Institution, DC for providing insights on the anatomy of beaked whales and B. Møhl and M. Wahlberg for providing helpful critique on earlier versions of the manuscript. We also thank the following for fieldwork assistance: N. Aguilar de Soto, M. Ballardini, A. Bocconcelli, J. F. Borsani, I. Ca-

varetta, T. Pusser, E. Revelli, and A. Sturlese. Tag development was assisted by T. Hurst, K. Shorter, K. Barton and funded in part by a Cecil H. and Ida M. Green Award and the US Office of Naval Research. WHOI fieldwork and tag development was funded by the National Oceanographic Partnership Program (NOPP), the Strategic Environmental Research and Development Program (SERDP) under Program No. CS-1188, and the Packard Foundation, and was supported by BluWest and the NATO Undersea Research Center. The whales were tagged under US National Marine Fisheries Service research Permit No, 981-1578-02 and 981-1707-00 to Peter L. Tyack and the research was approved by the Woods Hole Oceanographic Institution Animal Care and Use Committee. This is Contribution No 11270 from the Woods Hole Oceanographic Institution.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Au, W. W. L. (**1997**). ''Echolocation in dolphins with a dolphin-bat comparison,'' Bioacoustics **8**, 137–162.

Au, W. W. L., Floyd, R. W., Penner, R. J., and Murchison, E. (**1974**). ''Measurements of echolocation signals of the Atlantic bottlenose dolphin, *Tursiops truncatus* Montagu, in open waters,'' J. Acoust. Soc. Am. **56**, 1280–1290.

Au, W. W. L., Ford, J. K. B., Horne, J. K., and Newman Allman, K. A. (**2004**). ''Echolocation signals of free-ranging killer whales (*Orcinus orca*) and modelling of foraging for Chinook salmon (*Oncorhynchus tshawytscha*),'' J. Acoust. Soc. Am. **115**, 901–909.

Au, W. W. L., and Herzing, D. L. (**2003**). ''Echolocation signals of wild Atlantic spotted dolphin (*Stenella frontalis*),'' J. Acoust. Soc. Am. **113**, 598–604.

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). ''Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*),'' J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W. L., Moore, P. W., and Pawloski, D. (**1986**). ''Echolocation transmitting beam of the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **80**, 688–691.

Au, W. W. L., Pawloski, J. L., Nachtigall, P. E., Blonz, M., and Gisner, R. C. (**1995**). ''Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **98**, 51–59.

Brownlee, S. M. and Norris, K. S. (**1994**). ''The acoustic domain,'' in *The Hawaiian Spinner Dolphin*, edited by K. S. Norris, B. Würsig, R. S. Wells, and M. Würsig (University of California Press, Berkeley), pp. 161–185.

Caldwell, D. K., and Caldwell, M. C. (**1971**). ''Sounds produced by two rare cetaceans stranded in Florida,'' Cetology **4**, 1–6.

Dawson, S. M., Barlow, J., and Ljungblad, D. K. (**1998**). ''Sounds recorded from Baird's beaked whales *Berardius bairdii*.'' Marine Mammal Sci. **14**, 335–344.

Frantzis, A. (**1998**). ''Does acoustic testing strand whales?'' Nature (London) **392**, 29.

Frantzis, A., Goold, J. C., Sharsoulis, E. K., Taroudakis, M. I., and Kandia, V. (**2002**). ''Clicks from Cuvier's beaked whales, *Ziphius cavirostris* (L),'' J. Acoust. Soc. Am. **112**, 34–37.

Hooker, S. K., and Whitehead, H. (**2002**). ''Click characteristics of northern bottlenose whales (*Hyperoodon ampullatus*),'' Marine Mammal Sci. **18**, 69–80.

Jazwinski, A. H. (**1970**). *Stochastic Processing and Filtering Theory* (Academic, New York).

Jepson, P. D. *et al.* (**2003**). ''Gas-bubble lesions in stranded cetaceans,'' Nature (London) **425**, 575–576.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., Aguilar de Soto, N., and Tyack, P. L. (**2004**). ''Beaked whales echolocate on prey,'' Proc. R. Soc. London, Ser. B **271**, S383–S386.

Johnson, M., and Tyack, P. L. (**2003**). ''A digital acoustic recording tag for measuring the response of wild marine mammals to sound,'' IEEE J. Ocean. Eng. **28**, 3–12.

Lurton, X. (**2002**). *An Introduction to Underwater Acoustics* (Springer, London), 347 pp.

Lynn, S. K., and Reiss, D. L. (**1992**). ''Pulse sequence and whistle production by two captive beaked whales, *Mesoplodon* species,'' Marine Mammal Sci. **8**, 299–305.

Madsen, P. T., Wahlber, M., and Møhl, B. (**2002**). ''Male sperm whale (*Physeter macrocephalus*) acoustics in a high latitude habitat: implications for echolocation and communication,'' Behav. Ecol. Sociobiol. **53**, 31–41.

Madsen, P. T., Kerr, I., and Payne, R. (**2004**). ''Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales (*Pseudorca crassidens*) and Risso's dolphins (*Grampus griseus*),'' J. Exp. Biol. **207**, 1811–1823.

Marten, K. (**2000**). ''Ultrasonic analysis of pygmy sperm whale (*Kogia breviceps*) and Hubb's beaked whale (*Mesoplodon carlhubbsi*) clicks,'' Aquat. Mamm. **1**, 45–48.

Møhl, B., Surlykke, A., and Miller, L. A. (**1990**). ''High intensity narwhal click,'' in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 295–304.

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). ''The monopulsed nature of sperm whale clicks,'' J. Acoust. Soc. Am. **114**, 1143–1154.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648.

Rasmussen, M. H., Wahlberg, M., and Miller, L. (**2004**). ''Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhyncus albirostris*),'' J. Acoust. Soc. Am. **116**, 1826–1831.

Rice, D. W. (**1998**). *Marine Mammals of the World, Systematics and Distribution* (The Society for Marine Mammology, Allen Press, Lawrence).

Schotten, M., Au, W. W. L., Lammers, M. O., and Aubauer, R. (**2003**). ''Echolocation recordings and localization of wild spinner dolphins (*Stenella longirostris*) and pantropical spotted dolphins (*Stenella attenuata*) using a four-hydrophone array,'' in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. F. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 383–400.

Simmonds, M. P., and Lopez-Jurado, L. F. (**1991**). ''Whales and the military,'' Nature (London) **337**, 448.

Zimmer, W. M. X., Johnson, M. P., D'Amico, A., and Tyack, P. L. (**2003**). ''Combining data from a multi-sensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*),'' IEEE J. Ocean. Eng. **28**, 13–28.

Zimmer, W. M. X., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2005**). ''Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis,'' J. Acoust. Soc. Am. **117**, 1473–1485.

Invariance of evoked-potential echo-responses to target strength and distance in an echolocating false killer whale

Alexander Ya. Supin[a)]
Institute of Ecology and Evolution of the Russian Academy of Sciences, 33 Leninsky Prospect, 119071 Moscow, Russia

Paul E. Nachtigall,[b)] Whitlow W. L. Au,[c)] and Marlee Breese
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, Kaneohe, Hawaii 96744-1106

(Received 26 October 2004; revised 25 February 2005; accepted 25 March 2005)

Brain auditory evoked potentials (AEPs) were recorded in a false killer whale *Pseudorca crassidens* trained to accept suction-cup EEG electrodes and to detect targets by echolocation. AEP collection was triggered by echolocation pulses transmitted by the animal. The target strength varied from −22 to −40 dB; the distance varied from 1.5 to 6 m. All the records contained two AEP sets: the first one of a constant latency (transmission-related AEP) and a second one with a delay proportional to the distance (echo-related AEP). The amplitude of echo-related AEPs was almost independent of both target strength and distance, though combined variation of these two parameters resulted in echo intensity variation within a range of 42 dB. The amplitude of transmission-related AEPs was independent of distance but dependent on target strength: the less the target strength, the higher the amplitude. Recording of transmitted pulses has not shown their intensity dependence on target strength. It is supposed that the constancy of echo-related AEP results from variation of hearing sensitivity depending on the target strength and release of echo-related responses from masking by transmitted pulses depending on the distance.
[DOI: 10.1121/1.1914150]

PACS numbers: 43.80.Lb [JAS] Pages: 3928–3935

I. INTRODUCTION

The biosonar of odontocetes (dolphins and toothed whales) has been a subject of investigation over the past few decades (Nachtigall and Moore, 1988; Au, 1993), but many of the basic mechanisms underlying echolocation functioning remain uncertain. In particular, the problem of automatic gain control needs to be investigated in more detail. The target strength and distance of the targets inspected by the animal's sonar vary widely. Therefore, the echo level varies by many tens of dB. Successful performance of the sonar requires that its receiving part (the auditory system) is capable of analyzing echoes under these widely varying conditions. It is not known yet whether this performance is achieved by a kind of automatic gain control which keeps the response of the auditory system to the echo relatively constant, or if the auditory system is equally capable of analyzing the echo-signals at a wide variety of neuronal response levels.

The variation of the echo level may be decreased by variation of the transmitted pulse level. Experiments in the wild have shown that dolphins and whales vary the intensity of their echolocation pulses according to the distance to the target, roughly at a rate of 20 dB per distance decade (Rasmussen *et al.*, 2002; Au and Benoit-Bird, 2003; Au and Herzing, 2003; Au and Würsig, 2004; Au *et al.*, 2004). To some extent, this may compensate for the echo attenuation with distance, although the compensation may not be complete since, depending on the target size, the echo attenuation varies from 20 to 40 dB per distance decade. Apart from that, variation of the transmitted pulse intensity does not influence the echo-to-transmission ratio. If this ratio is very low (because of the low target strength and/or large distance), the echo may be masked by the much more intense preceding transmitted pulse. So, the question remains, whether or not other mechanisms of gain control may act in the odontocete's sonar, and if they do, which ones?

One way to answer this question is to record the brain auditory evoked potentials (AEPs) during natural echolocation in dolphins. AEPs indicate magnitude of the brain response to sounds. Recording AEPs during echolocation may show how the brain responds to both the emitted click and echo. Studies in a false killer whale (Supin *et al.*, 2003, 2004) have shown that this approach is feasible. During active echolocation experiments, a set of AEPs was recorded containing responses to both the transmitted sounds and to the echoes. This approach has revealed the presence of a self-regulating mechanism in the whale's auditory system. The echo-related AEP amplitude was virtually independent of the distance to the target when it varied from 1 to 8 m, although this resulted in echo-level variation as much as 36 dB (Supin *et al.*, 2004). It was suggested that this independence of the response to distance may have resulted from the interaction between the transmitted pulse and the echo: with the distance increase, the echo intensity decreases, but, because of increased delay, the echo releases from masking by

[a)]Electronic mail: alex_supin@sevin.ru
[b)]Electronic mail: nachtiga@hawaii.edu
[c)]Electronic mail: wau@hawaii.edu

 0001-4966/2005/117(6)/3928/8/$22.50

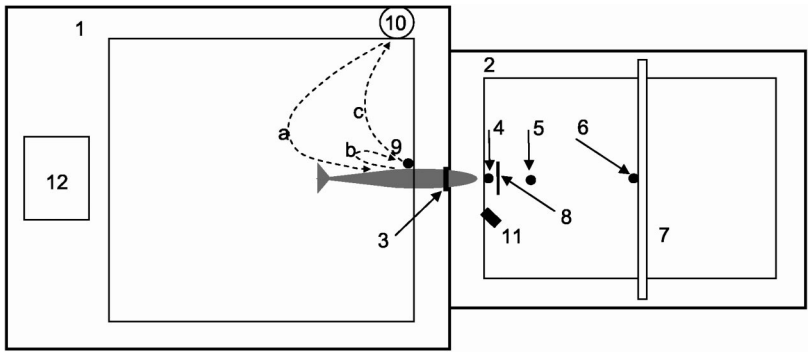


FIG. 1. Experimental conditions. Items 1–12 are designated in the text.

the preceding transmitted pulse; these two processes compensate one another keeping the echo-response almost constant. That study, however, was done with the use of only one target with a particular target strength. It left unanswered a few important questions, in particular:

(1) Does the odontocete's echolocation system use gain-control mechanisms compensating target strength as well as distance variations?
(2) Do these mechanisms, if they exist, involve control of the transmitted pulse intensity, or the control of hearing sensitivity, or both?

Therefore, the topics of the present study are as follows.

(1) To estimate how the auditory system responds to both transmitted clicks and echoes at various target strengths and distances, we recorded echolocation-related AEPs while varying both target strength and distance to the target.
(2) To estimate the role of transmitted click intensity in the automatic gain control, during a part of experiments, we recorded these outgoing echolocation clicks in parallel with the AEP recording.

II. MATERIALS AND METHODS

A. Subject and experimental conditions

The experiments were carried out in facilities of the Hawaii Institute of Marine Biology, Marine Mammal Research Program. The subject was a false killer whale *Pseudorca crassidens*, an approximately 30-year-old female kept in a wire-net enclosure in Kaneohe Bay, HI. The animal was trained to accept soft latex suction cups containing EEG electrodes to pick up the evoked-potentials, to ensonify and recognize targets by echolocation, and to report the target presence or absence using a go/no-go reporting paradigm.

The experimental facilities were laid out as follows (Fig. 1). The experimental enclosure was constructed of a floating pen frame (1), 8×10 m in size, supported by floats and bearing an enclosing wire net. This enclosure (the animal section) linked to a target section—another floating frame (2), 6×8 m in size that served to mount targets and did not bear net. The front part of this section was free of supporting floats and any other sound-reflecting objects in the water. In the net divider separating these two sections, there was an opening bounded by a hoop (3), 55 cm in diameter that served as a hoop station for the animal. In front of the hoop, a B&K 8103 hydrophone (4) was positioned to detect the emitted echolocation pulses and to trigger the acquisition system. Another B&K 8103 hydrophone (5) was positioned 1 m in front to record the echolocation pulses in free field. A target (6) was mounted on a wooden beam (7) fastened above the water surface across the target-section frame. The beam could be fixed at various distances from the animal. The target was hung from a thin monofilament line and could be pulled up out of water and lowered down into water. The hoop station (3), the hydrophones (4 and 5), and the lowered target (6) were in a longitudinal straight line; altogether at a depth of 80 cm. Behind the hydrophone (4), there was a movable baffle (8). When pulled up, this baffle screened the target section from the animal positioned in the hoop station; when it was lowered down, it opened the space in front of the animal. Near the hoop station, a response ball (9) was mounted above the water surface serving as a target-present response indicator. The trainer kept a position (10) to give instructions to the animal and to reward it with fish for correct responses. The animal's position in the stationing hoop was monitored through an underwater video camera (11). The electronic equipment and the operator were housed in a shack (12).

B. Experimental procedure

Each session included both target-present and target-absent trials assuring that the animal was indeed echolocating the target. The experimental procedure was as follows.

(i) Each session began with the trainer attaching suction-cup electrodes for AEP recording (see the following for details). (ii) The animal was sent to a hoop station (trace *a* in Fig. 1). During the animal positioning, the baffle [8 in Fig. 1(a)] screened the target from the animal. The target was either lowered down into water (a target-present trial) or pulled up out of water (a target-absent trial) in advance. (iii) As soon as the animal took the position in the hoop station, the baffle was lowered down thus opening the space in front of the animal. Immediately after that, the animal emitted a train of echolocation clicks, as a rule, 20–50 clicks in a train. (iv) When the target was present, the animal was required to signal its detection by leaving the hoop and touching the signal ball [trace *b* in Fig. 1(a)], then coming to the trainer for the fish reward (trace *c*). In the no-target trails, the animal was required to wait until it was signaled to leave the hoop (trace *b* was absent) and come for the fish reward (trace *c*).

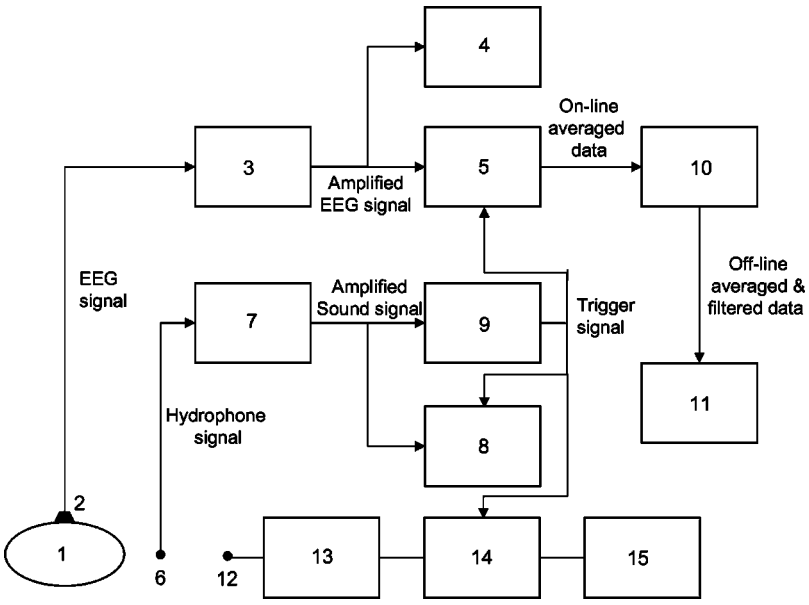


FIG. 2. Instrumentation. Items 1–15 are designated in the text.

Each session consisted of 30 trials, 20 target-present and 10 target-absent, randomly alternated. During the target-present trials, the AEP-acquisition system (see the following) was turned on as soon as the baffle was lowered, and it was kept on until the end of the echolocation click train. Thus, AEPs triggered by echolocation clicks were collected. In target-absent trials, AEPs were not collected because the animal emitted very few clicks and it was not possible to collect satisfactory AEP records. The purpose of the target-absent trials was to assure that the animal was functionally echolocating and reporting the presence and absence of targets.

C. Instrumentation

1. AEP recording

The AEP-recording equipment was designed as shown in Fig. 2. Brain potentials were picked up from the subject (1) by EEG electrodes (2) which were golden-plated disks 10 mm in diameter mounted within rubber suction cups 60 mm in diameter. The active electrode was attached with conductive gel at the dorsal head surface, at the midline, 5–7 cm behind the blowhole. The reference electrode was also attached along with conductive gel on the animal's back near the dorsal fin. Brain potentials were led by shielded cables to a balanced EEG amplifier (3) and amplified by 5×10^4 within a frequency range from 200 to 5000 Hz. The amplified signal was monitored by an oscilloscope Tektronix TDS1002 (4) and entered into a custom-made acquisition instrument (5) that contained a 12-bit analog-to-digital converter with a sampling rate of 20 kHz and a digital averager. The latter served to extract low-amplitude AEPs from background noise.

The averaging was triggered by sound clicks emitted by the animal. For that, sounds were picked up by a B&K 8103 hydrophone (6) positioned in front of the animal's head. The hydrophone signal was amplified by 40 dB by a custom-made sound amplifier (7). The amplified sound pulses were monitored by a digital-storage oscilloscope Tektronix TDS1002 (8) and entered a trigger device (9). Each time when a sound pulse appeared at the trigger-device input, it produced a standard pulse at its output. These standard pulses served to trigger the evoked-response averager (5); the triggering pulses were monitored with the same oscilloscope (8) as the sound pulses. After each click, a 15-ms long temporal window of evoked-potential recording was picked up, all the windows within the trial were averaged. The on-line averaged AEP records were stored in PC memory (10). The number of on-line averaged responses in a trial varied from 10 to 30, depending on the number of echolocation clicks picked up. This number was not sufficient to achieve good extraction of AEPs from noise. Therefore, all records obtained under similar conditions (one and the same target strength and distance) were then averaged off-line. The final AEP record was stored in PC memory (11).

2. Echolocation pulse recording

For echolocation pulse recording, they were picked up by a separate B&K 8103 hydrophone [(12) in Fig. 2], amplified by 12–18 dB by a custom-made amplifier (13) that had a gain-dependent upper frequency cutoff of approximately 1 MHz, digitized at a sampling rate of 1 MHz by a Measurement Computer PCI-DAS4020 acquisition board (14) and recorded in a PC memory (15). Echolocation pulses were collected using the same trigger pulses as for simultaneous AEP acquisition. Thus, those and only those echolocation pulses were collected which provoked the simultaneously collected AEPs. The echolocation signal data were analyzed with a custom QUICKBASIC 45 program. The parameters of interest included the source level, peak and center frequencies, 3-dB and rms bandwidths.

D. Targets

The targets used in active-echolocation experiments were hollow aluminum cylinders with an outer diameter of 38 mm (1.5 in.) and 25.5 mm (1 in.) inner diameter, axis vertical. Four targets were used, 180, 90, 45, and 23 mm long. Their target strengths were evaluated as -22 dB, -28,

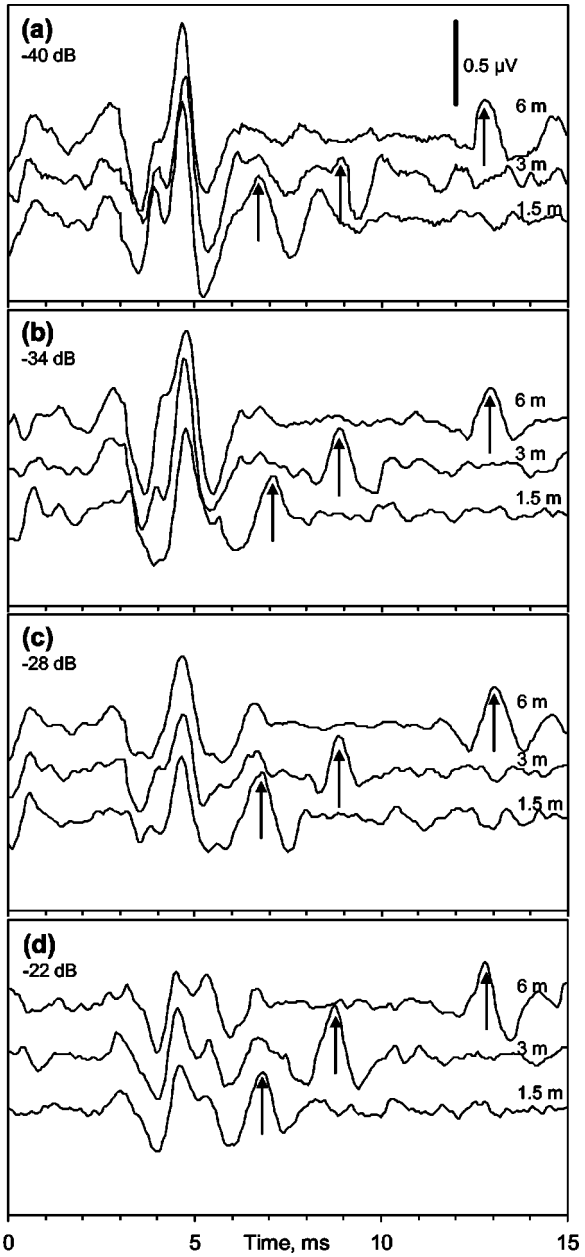


FIG. 3. Echolocation-related AEPs recorded at different target strengths [−40 to −22 dB, as designated at panels (a)–(d)] and different target distances (1.5–6 m, as designated near records in each panel). Arrows indicate the AEP complex considered as echo-related.

−34, and −40 dB, respectively. Cylindrical targets of fixed outer diameter and wall thickness were used so that varying the length could vary target strength without significantly changing the target frequency response.

III. RESULTS

A. AEP at various target strengths and distances

AEP records obtained at various target strengths and distances are summarized in Fig. 3. Each record was obtained by averaging 600–800 individual responses. Four panels, (a)–(d), present records obtained with four targets: 23, 45, 90, and 180 mm long with target strengths of −40, −34, −28, and −22 dB, respectively. Three traces in each of the panels present records at target distances of 6, 3, and 1.5 m, as indicated.

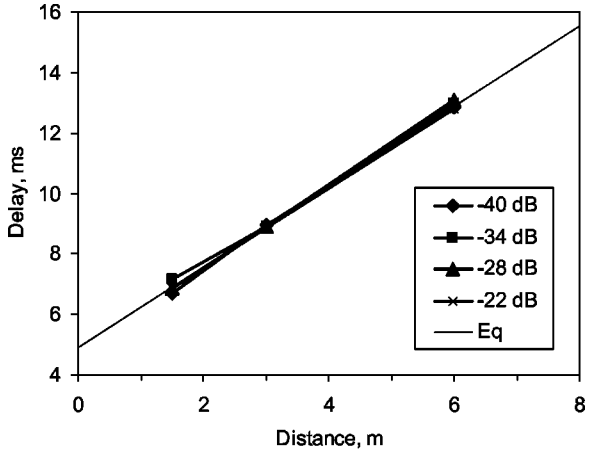


FIG. 4. Dependence of echo-related AEP latency on target distance, at four target strengths, −40 to −22 dB, as designated. Eq ■ is the approximating straight line drawn according to Eq. (1).

A common feature of all of the records was the presence of two AEP sets. The first one consisted of several alternative positive and negative waves; the latency of this AEP complex was one and the same in all records, independent of the target position: the first positive peak at 3.6–4 ms, the next negative peak at 4.6–4.8 ms, etc. The second AEP complex (marked by arrows in Fig. 3) was of somewhat simpler wave form (mostly positive–negative–positive), and its latency directly depended on the distance to the target: the longer the distance, the longer the latency.

The latency of the second AEP complex is plotted as a function of the target distance in Fig. 4. At all the target strengths, the plots fit a straight line. For comparison, a straight line described by

$$d = t + 1.33l \tag{1}$$

is shown in Fig. 5, where d is the delay (ms), l is the distance (m), and t is a constant (ms). The factor of 1.33 ms/m corresponds to the double-way delay-versus-distance dependence at the sound velocity of 1500 m/s. At $t = 4.9$ ms, the line almost coincides with the delay-versus-distance plots. We consider this coincidence as evidence that the second AEP complex is the response to the echo from the presented target. Therefore, this AEP complex is designated in the following as the echo response, or echo-AEP.

The first AEP complex independent of the target distance is considered as the response to the transmitted echolocation pulse. In the following it is designated as the transmission response, or transmission-AEP.

With this interpretation of the recorded AEPs, their following features may be pointed out as remarkable.

(1) The amplitude of the transmission-AEP was practically independent of target distance within the tested range from 1.5 to 6 m. However, the amplitude of this AEP was greatly dependent on target strength within the

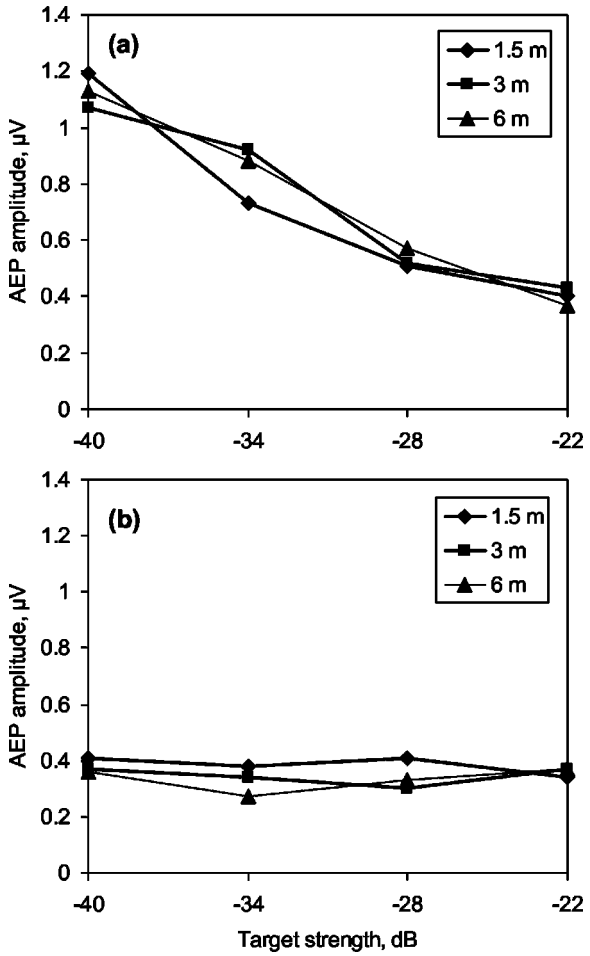


FIG. 5. (a) Transmission-related AEP amplitude dependence on target strength (the argument) and distance (the parameter). (b) The same for echo-related AEP.

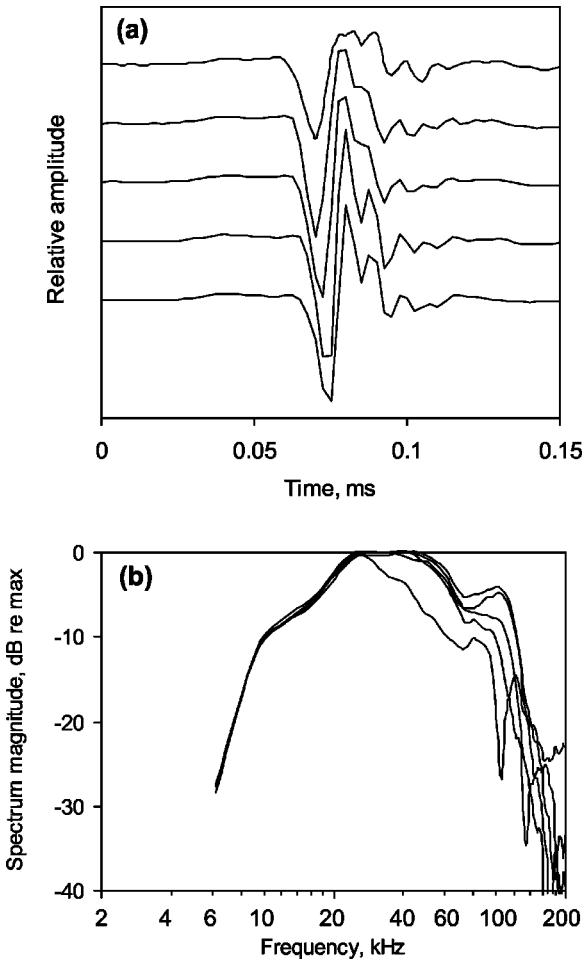


FIG. 6. Representative examples of transmitted echolocation pulse wave forms (a) and their spectra (b).

tested range of -22 to -40 dB: the more target strength, the less AEP amplitude [Fig. 5(a)].

(2) The amplitude of the echo-AEP was almost independent of both distance and target strength [Fig. 5(b)]. Although there were some differences between the records in the wave form and amplitude of the echo-AEPs, this difference was not systematic and looked more like random variation rather than a trend indicating dependence on target strength or distance.

B. Echolocation pulses at different target strengths and distances

Representative examples of echolocation pulses are presented in Fig. 6. Their spectra peaked at frequencies from 27 to 38 kHz. Echolocation pulse intensity varied significantly both within a train and from train to train. Their source levels were within a range from 165 to 210 dB re 1 μPa peak-to-peak, a majority (95%) being within a range of 175–200 dB. Figure 7 presents source level distributions of pulses collected at different target distances (from 1.5 to 6 m) at a target strength of -34 dB, and Fig. 8 presents the same for different target strengths (from -40 to -22 dB) at a distance of 3 m. Neglecting moderate data scatter, all the distributions look similar. Formally, almost all histograms presented in Figs. 7 and 8 were significantly different from one another by their means: differences varied from 1.6 dB [Figs. 8(a)–8(c)] to 7.5 dB [Figs. 7(a) and 7(b)] with standard errors from 0.21 to 0.36 dB. The only insignificant difference (0.4 dB with a standard error of 0.31 dB) was between the histograms in Figs. 7(b) and 7(c). However, there was no systematic dependence of the histogram means on either target strength or distance. Their means demonstrate very small (within a few dB) dependence on target distance [Fig. 9(a)] and no significant dependence on target strength [Fig. 9(b)].

IV. DISCUSSION

A. AEP independence of target distance

Independence of echo-AEP of target distance has been described previously (Supin *et al.*, 2004) at a target strength of -22 dB. Herein, the same regularity was confirmed for a certain range of target strengths, from -40 to -22 dB. Thus, this independence may be considered as a real feature of the subject's biosonar.

Among the possible mechanisms for the echo-response independence of distance, first of all the ''saturation'' mechanism must be considered. At a certain intensity of sound clicks, AEP amplitude may reach a maximal possible value, so that further intensity increase does not result in further response increase. In the odontocete's auditory system, this

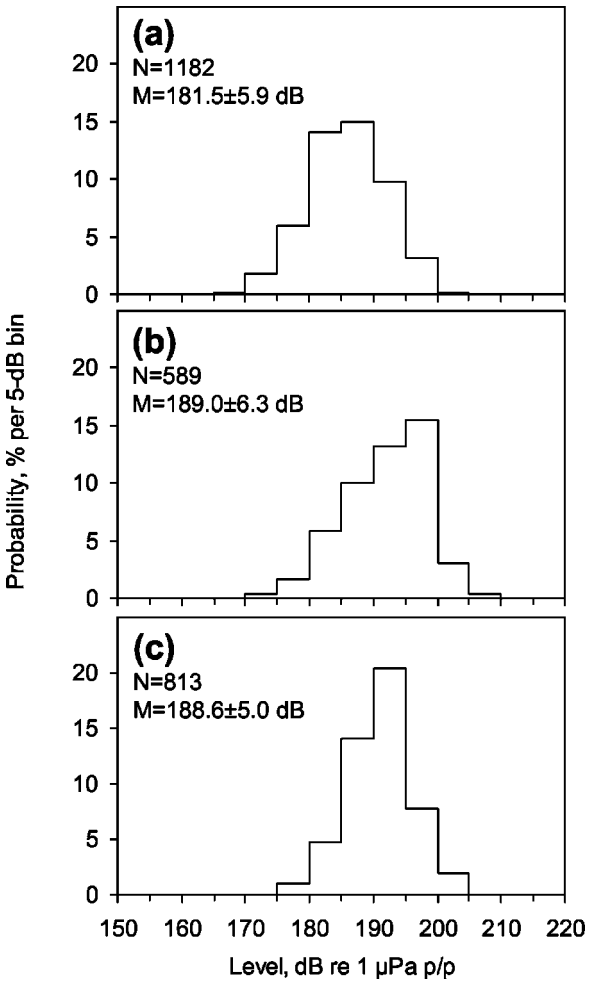


FIG. 7. Distributions of transmitted pulse source levels at different target distances: 1.5 m (a), 3 m (b), and 6 m (c). *N* is the number of collected pulses, *M* is the source level mean ± SD.

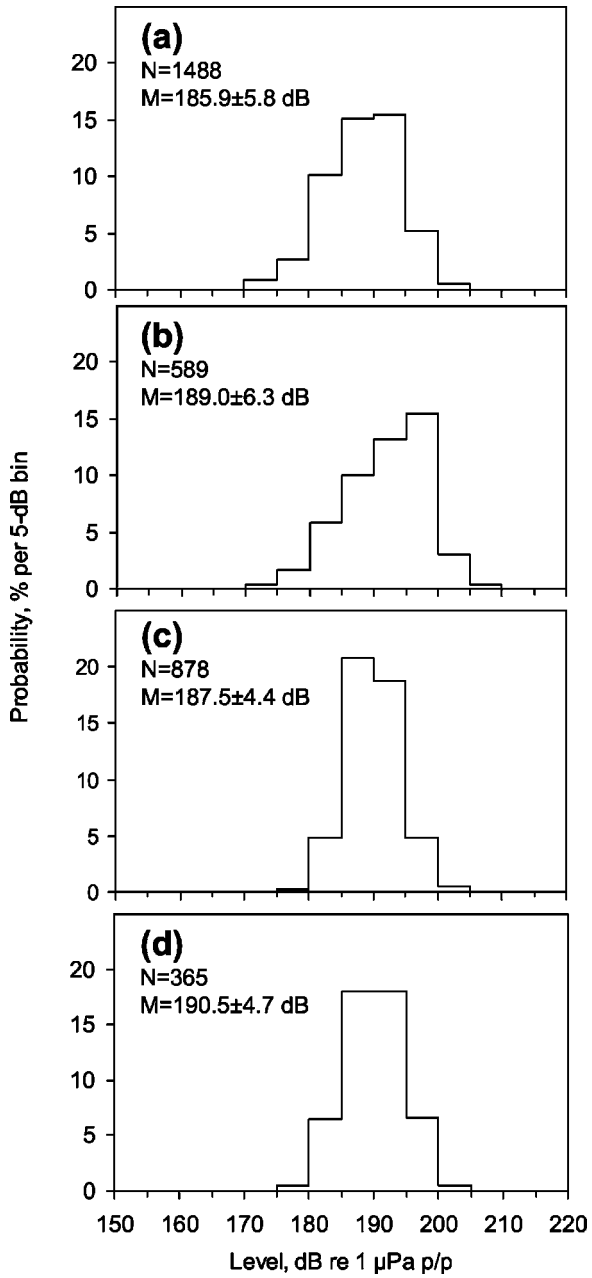


FIG. 8. Distributions of transmitted pulse source levels at different target strengths: −40 dB (a), −34 dB (b), −28 dB (c), and −22 dB (d). *N* is the number of collected pulses, *M* is the source level mean ± SD.

sort of saturation does appear, depending on the click spectrum, at levels from 70 to 80 dB above threshold (Popov and Supin, 1990) to 25–30 dB above threshold (Popov and Supin, 2001). Indeed, transmission-related AEP reached much higher amplitudes (more than 1 μV) than the echo-AEP (around 0.4 μV) recorded in the very same conditions, so the echo-AEP amplitude of 0.4 μV is well below the saturation level. Therefore, a kind of gain-control mechanism should be expected to keep the AEP amplitude independent of target distance.

One of the possible mechanisms of the gain control might be the variation of source level of the transmitted pulses. Although found in a number of field studies (see Sec. I), this mechanism was hardly involved in our experiments. Measurements of transmitted pulse level at different target distances demonstrated small and irregular variation of the mean level of only a few dB, whereas compensation of the distance change from 1.5 to 6 m required 24 dB. Independence of transmission-AEP amplitude on distance has directly shown that this small variation was negligible for the animal's auditory system. The ratio of the transmission- to echo-AEP remained constant when the target distance changed from 1.5 to 6 m which results in the echo attenuation change as much as by 24 dB. This means that the gain control is not a long-term process affecting both the transmission- and echo-responses but a short-term process affecting only the echo perception.

A hypothesis concerning the short-term gain control in the dolphin's auditory system has already been suggested (Supin *et al.*, 2004). It was supposed that the echo response is kept constant because of its delay-dependent releasing from masking by the previous louder transmitted pulse. With the target distance increase, the echo delay increases proportionally, which results in release from masking. The rate of this release in dolphins is 10–12 dB per delay doubling (Popov and Supin, 1990). At the same time, the echo intensity decreases with a rate of up to 12 dB per distance doubling. These two opposite processes may compensate one another, thus keeping the echo response almost constant. We

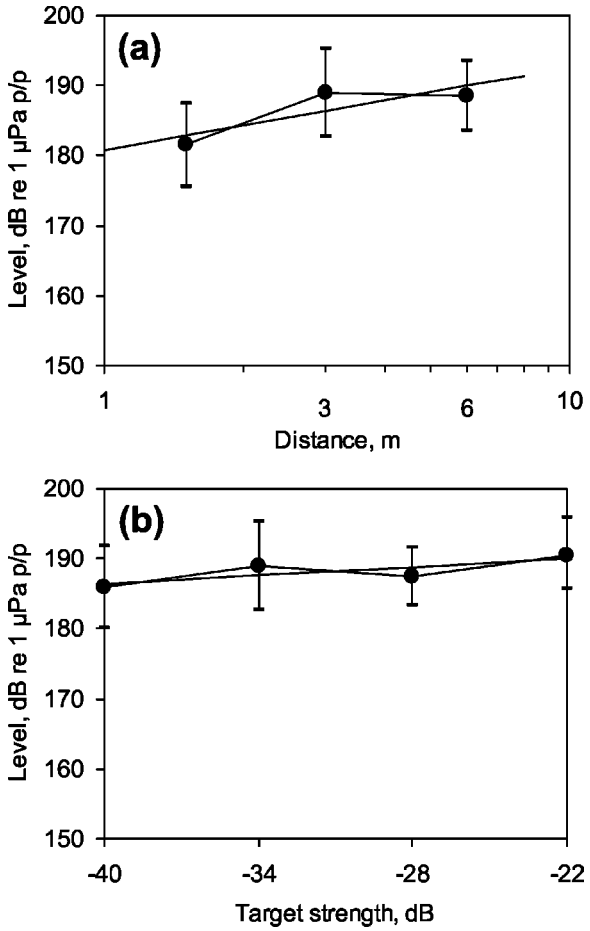


FIG. 9. Transmitted pulse source level dependence on target distance (a) and on target strength (b) (means ± SD). Straight lines are regression lines approximating the data.

may apply the same interpretation to the results obtained herein with a few different targets.

B. Echo-AEP independence on target strength

The results presented herein showed that *echo-AEPs* were independent both of target distance and of target strength that varied by as much as 18 dB. However, contrary to the target distance variation, the target strength variation strongly influenced the *transmission-AEP* amplitude, which increased almost three times when the target strength decreased from −22 to −40 dB. In other words, when the target strength decreased, the echo- to transmission-response ratio decreased correspondingly. It indicates the absence (or weak effectiveness) of a short-term mechanism influencing the echo-but not the transmission-AEP with the target strength variation. It is more likely that a long-term mechanism influences both of these responses.

Among such mechanisms, first of all the variation of the transmitted pulse intensity must be considered. Variation of transmitted pulse intensity (the lower the target strength, the higher the transmission intensity) could compensate the target strength variation without compensating the echo-to-transmission ratio. However, the direct measurement did not show any significant and systematic variation of the transmitted pulse depending on target strength. The overall trend of the plot presented in Fig. 9(b) is as small as +0.02 dB source level per 1 dB target strength, whereas compensation of target strength variation requires −1 dB/dB.

The only remaining explanation is variation of *hearing sensitivity*. Indeed, the response magnitude actually depends not on SPL stimulus intensity but on its sensation level (SL). If SL of the transmitted pulse changed without changing SPL, it indicates a change of sensitivity. Thus, we hypothesize that in our experiments the whale was capable of changing its hearing sensitivity in a manner compensating for the changes in the target strength: the lower the target strength, the higher the sensitivity. This mechanism could keep the echo response constant when the target strength varied, but it made the transmission response dependent on the target strength: at high target strength, the transmission response was low because of low sensitivity; at low target strength, it was high because of high sensitivity.

Because invasive investigations are not available in whales and dolphins, there is no definite data on the possible nature of the supposed regulation of hearing sensitivity. We may only hypothesize based on fundamental hearing mechanisms. It is commonly known that hearing sensitivity may be regulated at both conductive (the stapedial reflex) and sensorineural levels (adaptation). All these mechanisms are known as being provoked by acoustical stimulation itself, reducing the hearing sensitivity with increasing sound level. However, in experiments described herein, the level of transmitted echolocation pulses did not vary, thus the sensitivity was not controlled in the classical manner, i.e., by stimulus intensity. Is it possible that in odontocetes similar regulations of sensitivity can be triggered in another way: not by sounds themselves but by echolocation activity? We cannot exclude this possibility. But of course, without direct physiological investigations, this capability remains an intriguing hypothesis.

C. Data generalization

The above-presented data were obtained from only one subject of a species which has not been widely investigated. This brings up the question: How widely may these data, and their explanations, be expanded to: (1) other subjects of the same species and (2) other odontocete species?

It seems probable that odontocetes are capable of the use of different degrees and different mechanisms of automatic gain control, depending on the situation. In the above-described experiment, the subject did not use distance-dependent variation of the transmitted pulse level, whereas field studies in other species have demonstrated the use of this manner of gain control (see Sec. I). Does this difference depend on different conditions (laboratory-versus-field), or different target-distance ranges, or different species, or individual subject preferences? The answer can only be obtained by further investigations with more subjects, with varying conditions, different species, etc. But at least the data presented herein demonstrate that gain control within the auditory system is certainly one of the possible mechanisms of the odontocete's biosonar.

ACKNOWLEDGMENTS

The study was supported by the Office of Naval Research, Award No. N0014-98-1-687 administered by Robert Gisiner and Mardi Hastings (USA) and the Russian Ministry of Science and Technology, Award No. NSh.2152.2003.4 (Russia). Funding is gratefully acknowledged. Work was conducted under U.S. Marine Mammal Permit 978-1567 issued to Paul Nachtigall by the U.S. National Marine Fisheries Service Office of Protected Resources.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). ''Automatic gain control in the echolocation system of dolphins,'' Nature (London) **423**, 861–863.

Au, W. W. L., Ford, J. K. B., Horne, J. K., and Allman, K. A. N. (**2004**). ''Echolocation signals of free-ranging killer whales (Orcinus orca) and modelling of foraging for chinook salmon (Oncorhynchus tshawytscha),'' J. Acoust. Soc. Am. **115**, 901–909.

Au, W. W. L., and Herzing, D. L. (**2003**). ''Echolocation signals of wild Atlantic spotted dolphin (Stenella frontalis),'' J. Acoust. Soc. Am. **113**, 598–604.

Au, W. W. L., and Würsig, B. (**2004**). ''Echolocation signals of dusky dolphins (Lagenorhynchus obscurus) in Kaikoura, New Zealand,'' J. Acoust. Soc. Am. **115**, 2307–2313.

Nachtigall, P. E., and Moore, P. W. B. (**1988**) *Animal Sonar: Processes and Performance* (Plenum, New York).

Popov, V. V., and Supin, A. Ya. (**1990**). ''Auditory brain stem ponses in characterization of dolphin hearing,'' J. Comp. Physiol., A **166**, 385–393.

Popov, V. V., and Supin, A. Ya. (**2001**). ''Contribution of various frequency bands to ABR in dolphins,'' Hear. Res. **151**, 250–260.

Rasmussen, M. H., Miller, L. A., and Au, W. W. L. (**2002**). ''Source levels of clicks from free-ranging white beaked dolphins (Lagenorhinchus albirostris Gray 1846) recorded in Icelandic waters,'' J. Acoust. Soc. Am. **111**, 1122–1125.

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2004**). ''The interaction of outgoing echolocation pulses and echoes in the false killer whale's auditory system: Evoked-potential study,'' J. Acoust. Soc. Am. **115**, 3218–3225.

Supin, A. Ya., Nachtigall, P. E., Pawloski, J., and Au, W. W. L. (**2003**). ''Evoked potential recording during echolocation in a false killer whale *Pseudorca crassidens*,'' J. Acoust. Soc. Am. **113**, 2408–2411.

Supin, A. Ya., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer, Dordrecht).

Pure tone audiograms and possible aminoglycoside-induced hearing loss in belugas (*Delphinapterus leucas*)

James J. Finneran[a] and Donald A. Carder
U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152-5001

Randall Dear
Science Applications International Corporation, Maritime Services Division, 3990 Old Town Avenue, Suite 105A, San Diego, California 92110

Traci Belting
Point Defiance Zoo and Aquarium, 5400 North Pearl Street, Tacoma, Washington 98466

Jim McBain and Les Dalton
Sea World, 500 Sea World Drive, San Diego, California 92109

Sam H. Ridgway
U.C. Veterinary Medical Center—San Diego, Department of Pathology, School of Medicine, University of California, La Jolla, California 92093-0612

(Received 7 October 2005; revised 10 February 2005; accepted 22 February 2005)

A behavioral response paradigm was used to measure pure-tone hearing sensitivities in two belugas (*Delphinapterus leucas*). Tests were conducted over a 20-month period at the Point Defiance Zoo and Aquarium, in Tacoma, WA. Subjects were two males, aged 8–10 and 9–11 during the course of the study. Subjects were born in an oceanarium and had been housed together for all of their lives. Hearing thresholds were measured using a modified up/down staircase procedure and acoustic response paradigm where subjects were trained to produce audible responses to test tones and to remain quiet otherwise. Test frequencies ranged from approximately 2 to 130 kHz. Best sensitivities ranged from approximately 40 to 50 dB re 1 μPa at 50–80 kHz and 30–35 kHz for the two subjects. Although both subjects possessed traditional ''U-shaped'' mammalian audiograms, one subject exhibited significant high-frequency hearing loss above 37 kHz compared to previously published data for belugas. Hearing loss in this subject was estimated to approach 90 dB for frequencies above 50 kHz. Similar ages, ancestry, and environmental conditions between subjects, but a history of ototoxic drug administration in only one subject, suggest that the observed hearing loss was a result of the aminoglycoside antibiotic amikacin.
[DOI: 10.1121/1.1893354]

PACS numbers: 43.80.Lb [WA] Pages: 3936–3943

I. INTRODUCTION

The first cetacean audiogram was obtained by Johnson (1966, 1967), who measured pure-tone thresholds in a trained bottlenose dolphin. Since that time, researchers have investigated hearing sensitivity, frequency selectivity, masking, auditory fatigue, temporal integration, and localization in dolphins and other marine mammal species (review Nachtigall, 1986; Johnson, 1986; Au, 1993; Nachtigall *et al.*, 2000). The majority of studies on the auditory capabilities of marine mammals have used psychophysical or behavioral response paradigms similar to that employed by Johnson (1966, 1967). In the behavioral method, the subject is trained to give a specific response to a particular acoustic stimulus and to withhold the response (or provide an alternate response) in the absence of the stimulus. Behavioral techniques allow direct measurements of hearing sensitivity and are generally considered the ''standard'' to which other sensitivity measures (e.g., electrophysiological measures) are compared (e.g., Szymanski *et al.*, 1999).

Behavioral methods are limited, however, by the difficulty and costs involved with training marine mammals to participate in hearing tests. Most marine mammal psychoacoustic studies have used one or two experimental subjects (Green *et al.*, 1994). Little attention has been given to replicating earlier work with additional subjects. The small number of individuals for whom data are available has resulted in lingering questions regarding ''normal'' hearing for marine mammal species, typical intraspecific variability, and typical hearing loss for different ages and genders.

In this paper we report behavioral audiograms for two belugas (*Delphinapterus leucas*). These data augment the beluga hearing threshold data presented by White *et al.* (1978), Awbrey *et al.* (1988), Klishin *et al.* (2000), and Ridgway *et al.* (2001). Taken together, these data allow estimates to be made of ''normal'' hearing sensitivity in belugas. The data from the present study were also unique because they

[a]Electronic mail: james.finneran@navy.mil

revealed significant hearing loss in a subject previously treated with aminoglycoside antibiotics.

II. METHODS

A. Subjects

Test subjects were two male belugas: Beethoven (9–11 years old, mass approximately 640 kg) and Turner (8–10 years, approximately 590 kg). Both subjects were born in an oceanarium, had the same father, and had been housed together since shortly after Turner was born. Neither subject had any previous experience with hearing tests or other psychophysical test procedures.

The health of the subjects was ascertained through periodic medical examinations by veterinarians. Subjects were healthy during the course of the study, with the exception that from June 2002 until the end of the study Turner was treated off and on for glomerulonephritis as evidenced by periodic hemituria. No ototoxic drugs were used during these periods of treatment. Beethoven required no medical treatment during the study period.

Tests were conducted according to a protocol approved by the Institutional Animal Care and Use Committees at the Space and Naval Warfare Systems Center, San Diego and the Point Defiance Zoo and Aquarium. The described experiments were conducted in accordance with the Acoustical Society of America's *Guiding Principles in the Care and Use of Animals* and followed all applicable U.S. Department of Defense guidelines.

B. Experimental apparatus

Figure 1(a) shows the test site located in the ''Rocky Shores'' beluga habitat at the Point Defiance Zoo and Aquarium, in Tacoma, WA. The exhibit contained approximately 1150 m^3 (304 000 gal) of filtered, ozonated seawater within a large main pool, a shallow connection channel, and an off-exhibit holding pool. The holding pool, which was 9 m in diameter with a depth of 2.7 m, was used to separate the subjects so each could be tested independently. The hearing tests were conducted in the main pool. The main pool was irregularly shaped, roughly 15×20 m, with a surface of sprayed gunite and varying bottom depth and topography. The maximum depth was 4.4 m, sloping up to 1.5 m at the entrance to the channel. The average depth was about 3 m. The main pool volume was 920 m^3 (243 000 gal) and the surface area was 30 m^2.

The trainer was positioned near a shallow beaching area at the northwest edge of the pool. The test apparatus was located along the south side of the pool, near an approximately 10-m-long underwater viewing window built into the southwest wall of the pool. A personal computer (PC), video monitor, and other electronics were housed in a small enclosure located in the underwater viewing area.

The test apparatus [Fig. 1(b)] consisted of a submerged polyvinyl chloride (PVC) frame containing an underwater sound projector (ITC 10001, ITC 1032, or ITC 1042, depending on the test frequency), receiving hydrophone (B&K 8105), video camera, and a neoprene-covered plastic ''biteplate.'' Subjects were trained to dive underwater and position

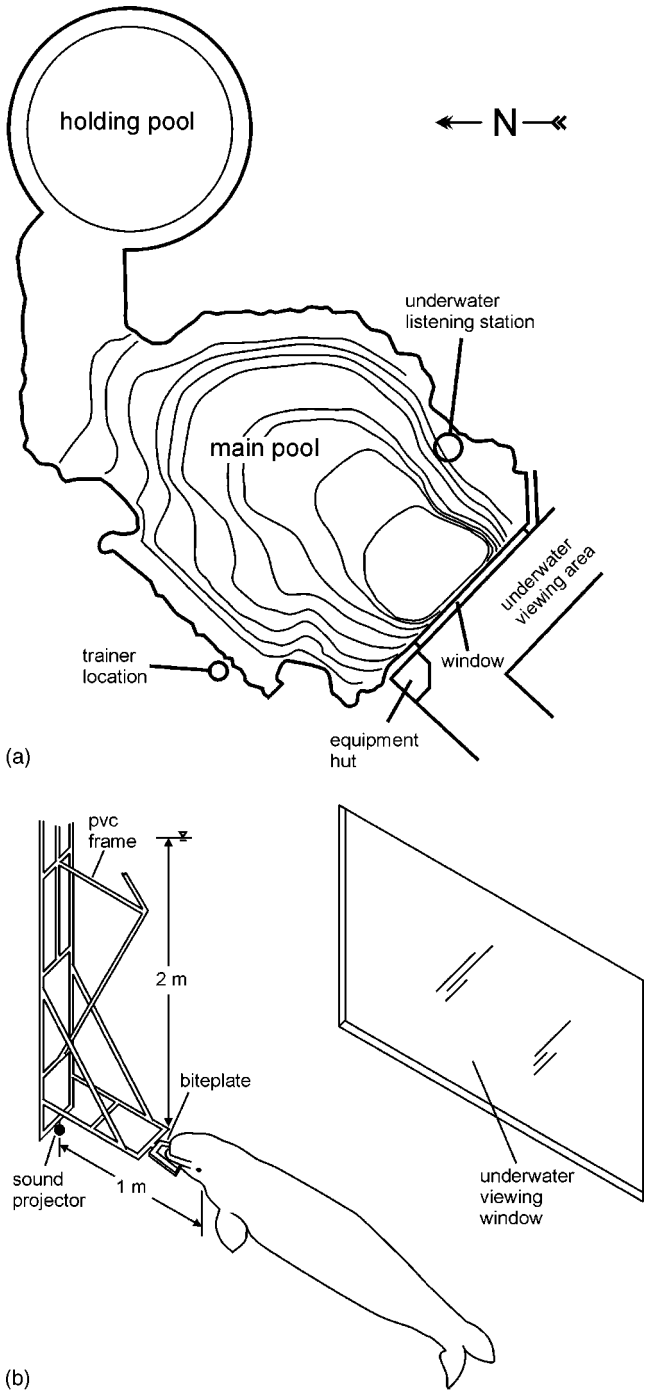


FIG. 1. (a) Top-view schematic of the beluga habitat at the Point Defiance Zoo and Aquarium showing the main pool and holding pool. Locations of the test apparatus and trainer positions are marked. Depth contours are approximate. (b) Schematic of the test apparatus with subject positioned on the biteplate.

themselves on the biteplate, which put their head in a fixed, repeatable position with respect to the sound projector and receiving hydrophone. Water depth at the apparatus was approximately 4 m. The depth of the sound projector and subject was approximately 2 m. The sound projector was positioned so that the distance between the projector and the subject's ears was approximately 1 m.

Hearing test tones were generated by a multifunction data acquisition board (National Instruments PCI-MIO-16E-1) residing within the PC. The generated tones were

attenuated (HP 355D or TDT PA-5), filtered (Ithaco 4302), and amplified (BGW PS2), before being input to the sound projector: ITC 1001 (2–20 kHz), ITC 1032 (12–50 kHz), or ITC 1042 (18–130 kHz). Tones were 500 ms in duration, including a 50-ms linear rise and fall time. Sound levels presented to each subject were calibrated before and after each session, without the subject present, with the receiving hydrophone located at a position estimated to lie on the subject's midline at the location of the ears. The hydrophone output was amplified (B&K 2635) and filtered (Ithaco 4302) before being digitized by the PCI-MIO-16E-1. During the hearing tests the hydrophone was positioned above and in front of the subject and used to monitor the sound in the water, including the hearing test tones and any sounds produced by the subject. Custom software (Finneran, 2003) was used to calibrate the sound system, control the hearing test, and analyze the resulting data.

C. Procedure

1. Hearing test

Hearing test sessions were conducted once or twice a day with each subject. Each session produced a single threshold estimate at a particular frequency. Session durations were approximately 15 to 20 min. The hearing test procedure was based on the method of free response (MFR) (Egan *et al.*, 1961). Similar test methods were used by Finneran *et al.* (2000, 2002a, b, c, 2003) and Schlundt *et al.* (2000) to measure marine mammal hearing thresholds.

Each hearing test session was divided into a number of observation periods, referred to here as ''dives.'' During each dive, the subject was instructed to swim to the test apparatus, submerge, and position at the underwater biteplate. A variable number of trials was then presented. To conclude the dive, the trainer sounded an underwater buzzer and signaled the subject to return for fish reward.

Each trial was 2 s in duration. The intertrial interval (defined from the start of one trial to the start of the next trial) was randomized between 4 and 7 s. Fifty percent of the trials (determined from a Gellerman series) contained a hearing test tone, 50% were signal-absent or ''catch'' trials. Tone trials contained a 500-ms duration pure tone at the test frequency. The tone onset coincided with the trial start. Subjects were trained to produce a specific audible response to each hearing test tone and to stay quiet otherwise. The time period between 0.1 and 2.0 s immediately following each tone start was designated as the ''hit interval.'' Responses occurring within a hit interval (following a tone) were recorded as ''hits.'' No response to a tone trial was classified as a ''miss.'' Catch trials were identical to tone trials except that the PC-generated tone amplitude was set to 0 V. Subjects were not aware of the trial start times, thus the catch trials were essentially extensions of the randomized intertrial interval from the proceeding tone and functioned as equipment catch trials (to ensure that the generation/recording process was not producing acoustic artifacts audible to the subjects). Whistle responses to catch trials were recorded as ''false alarms'' (whistle responses by a subject outside of any trial period were also recorded—see below). No response to a catch trial was recorded as a ''correct rejection.'' Tone amplitudes were adjusted using a modified up/down staircase procedure (e.g., Cornsweet, 1962): the amplitude was decreased 2 dB after each hit and increased 2 dB after each miss. Hearing thresholds for a single session were estimated from the mean sound pressure of 10 hit-miss/miss-hit reversal points collected within that session.

The trainer and computer operator monitored the sound in the water for any responses by the subject. A small LCD was used to display trial parameters (e.g., stimulus level, tone or catch trial, dive time) for the trainer. The display was updated just before the start of a trial. The number of trials per dive was randomized within the following guidelines: Dives were ended only after correct responses. An attempt was made to reinforce responses to low-level tones (i.e., at a lower level than any previously responded to). The first hit following several misses was generally not reinforced. The dive times were normally kept under 2 min. The amount of fish reward was scaled to the performance of the subject during the dive (e.g., more reinforcement was given for longer dives and/or responding to low-level tones).

2. False alarms

Previous studies of marine mammal audition have demonstrated the importance of the subject's motivational state and response bias (e.g., Schusterman *et al.*, 1975). In the present study, the response bias was assessed using two techniques. In the first, more traditional method, the false alarm rate R_{FA} was defined as

$$R_{\mathrm{FA}}=\frac{N_{\mathrm{FA}}}{N}, \tag{1}$$

where N_{FA} is the number of false alarms and N is the total number of trials. Miller (1969) presented an alternate method to assess response bias in the MFR:

$$r_{\mathrm{FA}}=\left(\frac{n_{\mathrm{FA}}}{T-N_T T_1}\right)T_1, \tag{2}$$

where r_{FA} is the false alarm rate, n_{FA} is the total number of whistle responses occurring outside of a hit interval, T is the total amount of time the subject spent on the biteplate, N_T is the number of tones presented, and T_1 is the hit interval duration. The term in parentheses is the number of ''false positive'' responses divided by the total amount of time during which the subject was on the station with no hit interval present. This term is multiplied by T_1 to obtain a dimensionless quantity. For the MFR, r_{FA} is analogous to R_{FA}; however, this study employed a modified version of the MFR where the intertrial interval was constrained between 4 and 7 s—it was not a Poisson distribution. For this reason, we report both false alarm measures. In most sessions, there were few responses outside of any response interval, so the two measures are nearly proportional.

3. Threshold estimates

Hearing thresholds were measured over a 20-month period. Thresholds were measured for Beethoven at 29 frequencies between 2 and 130 kHz. For Turner, thresholds

were measured at 28 frequencies between 2 and 100 kHz. Above 100 kHz, Turner did not respond to tones with an SPL of 160 dB *re* 1 μPa and was therefore not tested at these frequencies.

Test frequencies were separated into three overlapping groups, dictated by the usable ranges of the three available sound projectors: low (2–20 kHz), mid (12–50 kHz), or high (18–130 kHz). Frequencies belonging to more than one group were tested with multiple projectors to ensure consistency in the thresholds despite changes in the sound source. Testing began at the low frequencies, then progressed to the mid and high frequencies. Within each group, the frequency was varied from day to day. After several months, testing shifted back to the low frequency group and the process was repeated for most frequencies, so thresholds were obtained several months apart.

Each frequency was tested at least three times; most frequencies were tested five or more times. Each test yielded an independent threshold based on ten reversals. The number of times a particular frequency was tested depended, in part, on the variability of the threshold measurements at a single frequency and between nearby frequencies. Additional tests were conducted at frequencies where measurements were highly variable and at frequencies where thresholds showed large differences compared to neighboring frequencies. The threshold and false alarm data at each frequency were used to calculate the mean and standard deviations for the threshold and false alarm rates as functions of the sound frequency.

III. RESULTS

Table I and Fig. 2 present the hearing thresholds, R_{FA}, and r_{FA} values for Beethoven. The symbols in Fig. 2 indicate the mean values; the error bars represent the 95% confidence intervals. Figure 2(a) includes a representative sample of the mean ambient noise spectral density level (in dB *re* 1 μPa^2/Hz) measured in the test pool. Above approximately 20 kHz ambient noise levels were below the self-noise of the measuring hydrophone and amplifier (B&K 8105 and B&K 2635). Table I includes the number of measurements conducted at each frequency (n). Table II and Fig. 3 present analogous data for Turner.

Audiograms for both subjects have the ''U-shape'' typically seen in mammals. Beethoven had best sensitivity from approximately 50 to 80 kHz and functional hearing (defined here as thresholds $<$120 dB *re* 1 μPa) above 100 kHz. Notches and peaks in sensitivity were observed at 20 and 50 kHz, respectively. Turner had best sensitivity from 30 to 35 kHz and functional hearing up to only about 50 kHz. False alarm rates in both subjects averaged near 5% and 1% for R_{FA} and r_{FA}, respectively. False alarm rates were variable from session to session, leading to relatively high standard deviations. There were no substantial differences in false alarm rates with frequency in either subject.

Figure 4 compares the data from the present study to beluga hearing thresholds previously measured by White *et al.* (1978), Awbrey *et al.* (1988), Klishin *et al.* (2000), and Ridgway *et al.* (2001). The adult male tested by Awbrey *et al.* was one of the subjects tested by White *et al.* White *et al.*, Awbrey *et al.*, and Ridgway *et al.* used behavioral methods; Klishin *et al.* used an electrophysiological technique. Ridgway *et al.* conducted measurements in the open ocean at 5-m depth; the other data were obtained in pools with depths from about 1.5 to 4 m. Klishin *et al.* lowered water depth to 40 cm during measurements, so that the electrodes remained above the waterline.

TABLE I. Hearing thresholds and false alarm rates for Beethoven. SD =standard deviation. n=number of independent threshold estimates (each based on ten reversals).

	Threshold		R_{FA}		r_{FA}		
Frequency (kHz)	Mean (dB *re* 1 μPa)	SD (dB)	Mean (%)	SD (%)	Mean (%)	SD (%)	n
2	89	5.3	1.6	3.7	0.3	0.6	12
4	82	5.4	5.7	6.5	1.4	1.6	13
5	77	2.1	5.4	7.3	1.7	1.3	6
7	67	3.6	3.8	5.0	1.6	2.0	6
8	67	1.8	2.3	3.6	1.6	1.8	6
10	68	3.3	3.1	5.3	1.1	2.0	10
12	67	4.4	3.5	3.3	0.7	0.7	7
14	62	4.1	4.6	7.4	0.7	1.2	7
15	65	1.7	13.0	2.0	2.2	0.4	4
18	61	3.3	10.2	7.6	1.7	1.3	4
20	77	4.3	3.8	6.9	0.6	1.1	16
25	61	6.8	4.3	4.2	0.8	0.8	7
30	60	4.5	7.6	8.8	1.0	1.1	5
40	57	2.1	3.0	5.9	0.5	1.0	7
45	59	1.0	5.8	6.3	1.0	1.1	3
50	43	2.4	2.6	3.6	0.5	0.6	5
55	55	4.8	1.8	3.6	0.3	0.6	4
60	53	1.2	7.2	7.7	1.2	1.2	3
70	53	4.6	9.8	6.1	1.7	1.2	4
80	56	1.9	7.1	8.2	1.3	1.5	4
90	59	7.7	5.6	6.2	0.9	1.1	13
91	72	8.9	1.1	2.5	0.2	0.5	5
95	75	2.9	2.0	3.4	0.4	0.8	3
100	74	4.0	6.5	5.9	1.1	1.0	14
110	79	3.9	13.4	8.9	2.5	2.0	7
115	86	3.1	2.6	4.4	0.5	0.9	3
117	92	1.0	2.1	3.6	0.4	0.7	3
120	101	2.3	2.3	3.6	0.4	0.6	6
130	103	2.7	5.1	8.7	0.8	1.4	5

At the lower frequencies (below 10 kHz), the data from the present study are consistent with the White *et al.* and Awbrey *et al.* data, which were also obtained in pools. Differences in ambient noise levels may explain the relatively low thresholds obtained by Ridgway *et al.* at these frequencies. At the higher frequencies, Beethoven's thresholds are close to those of the subjects tested by Ridgway *et al.* and the male tested by White *et al.* The relatively high thresholds reported by Klishin *et al.* may be the result of the evoked potential methodology, the very shallow water, which could have created problems in accurately assessing the received SPL, or the short duration stimuli (about 20 ms), which may have resulted in higher thresholds due to temporal integration (Johnson, 1968). Both subjects tested by White *et al.* exhibited peaks and notches in sensitivity similar to those seen in Beethoven. The upper cutoff frequency in Beethoven is close to that observed in the other subjects. Turner had lower sensitivity at the higher frequencies compared to other belugas. The dramatic increase in Turner's thresholds above

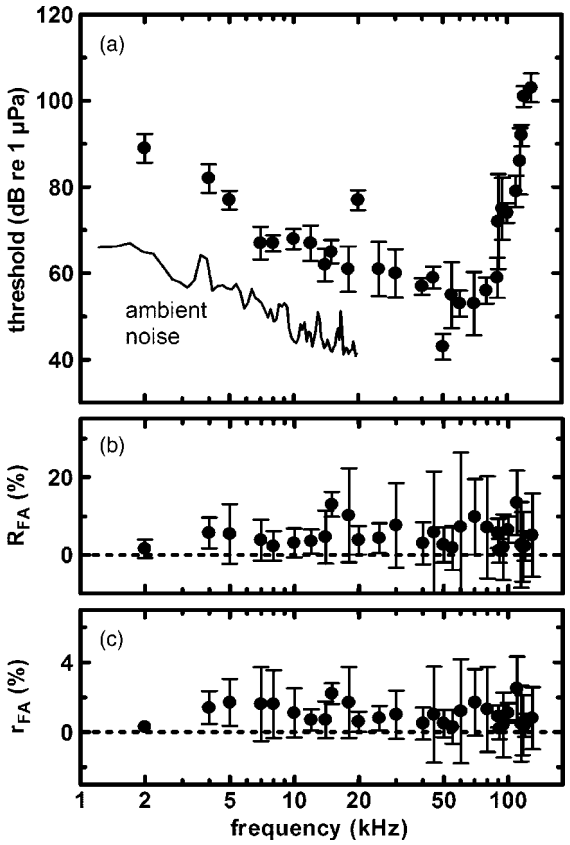


FIG. 2. (a) Hearing thresholds, (b) R_{FA}, and (c) r_{FA} values as functions of frequency for Beethoven. The symbols indicate mean values, the error bars represent the 95% confidence intervals.

TABLE II. Hearing thresholds and false alarm rates for Turner. SD =standard deviation. n=number of independent threshold estimates (each based on ten reversals).

	Threshold		R_{FA}		r_{FA}		
Frequency (kHz)	Mean (dB *re* 1 μPa)	SD (dB)	Mean (%)	SD (%)	Mean (%)	SD (%)	n
2	93	3.5	1.3	2.8	1.3	0.6	5
4	81	4.8	4.5	4.5	5.7	2.1	11
5	76	1.3	0.9	2.4	2.3	0.9	7
7	72	2.4	2.6	4.3	5.2	1.9	6
8	68	3.2	1.7	2.9	1.1	0.5	7
10	64	2.2	2.9	3.4	1.7	0.7	8
12	72	5.1	4.9	6.9	4.6	2.0	7
14	67	1.5	0.0	0.0	0.0	0.0	4
15	68	2.4	6.3	7.7	2.9	1.4	4
20	60	2.1	1.7	3.7	1.8	0.7	10
25	54	1.3	8.8	10.4	3.6	1.9	4
30	63	5.3	4.5	6.6	3.7	1.2	11
32	52	4.5	4.1	4.9	1.6	0.8	4
34	52	1.7	4.2	8.3	2.9	1.5	4
35	55	4.0	8.5	9.1	3.9	1.7	4
37	65	5.5	5.9	6.1	2.8	1.2	5
38	67	3.5	3.2	5.5	1.7	1.0	3
40	79	3.3	3.6	5.6	1.9	1.0	6
42	84	3.6	10.9	3.9	3.0	0.7	5
44	102	3.5	4.1	3.7	1.3	0.7	3
46	114	3.6	9.7	0.5	1.9	0.3	3
48	133	5.1	8.2	4.3	2.4	0.8	6
50	142	2.7	8.6	6.7	3.7	1.3	6
60	140	3.3	10.8	7.9	3.2	1.4	4
70	146	5.6	8.6	8.3	3.1	1.6	3
80	143	3.0	6.7	6.7	2.4	1.2	3
90	151	3.5	7.6	2.3	1.9	0.5	3
100	161	2.6	2.1	3.6	1.2	0.7	3

35 kHz is unique among the individual belugas for which data exist.

Although each of the previous studies of beluga hearing utilized only one to three individual subjects, pooling these data allows a larger sample size to be obtained. From this, estimates may be made of what typical means and variations may be in beluga hearing thresholds. To accomplish this, the data from Fig. 4 were collapsed at each frequency to create a single composite audiogram. The following restrictions were used when pooling the data: Only behavioral psychoacoustic data was used, so the data from Klishin *et al.* (2000) were excluded. Turner's thresholds were also excluded. The individual subject tested by both Awbrey *et al.* (1988) and White *et al.* (1978) was only included once (the White *et al.* data were used). Linear interpolation was used to estimate thresholds at intermediate frequencies. Figure 5 shows the resulting composite beluga audiogram. The solid and dashed lines indicate the mean and the mean±one standard deviation, respectively.

Figure 5 allows comparisons to be made between Turner's hearing thresholds and ''normal'' thresholds for belugas. Figure 6 shows the differences between Turner's hearing thresholds and the mean values from Fig. 5 and represents the estimated hearing loss for Turner at each frequency.

IV. DISCUSSION

The hearing thresholds presented in Figs. 2 and 3 exhibit the typical mammalian ''U-shape;'' however, the large differences between Turner and Beethoven's thresholds raise numerous questions. The large differences between thresholds are especially interesting considering the identical environments and test conditions to which the animals were subjected. Hearing loss up to 15 to 25 dB is often considered ''normal'' for humans and to represent no impairment or handicap (Davis and Silverman, 1978; Kinsler *et al.*, 1982; Glorig, 1988; ASLHA, 2004). According to this definition, Turner's thresholds would be considered normal below about 37 kHz. At 50 kHz and above Turner's hearing loss is approximately 90 dB, which would be considered ''severe'' hearing loss in humans (Davis and Silverman, 1978; Kinsler *et al.*, 1982; ASLHA, 2004).

Hearing loss in mammals can be caused by a variety of factors, including aging, exposure to high intensity sound, exposure to ototoxic drugs, or congenital factors (review Pickles, 1988; Yost, 1994). Ridgway and Carder (1993, 1997) reported hearing deficits in three male dolphins (23, 26, and 34 years) and one female dolphin (33 years). Brill *et al.* (2001) also reported hearing loss above 55 kHz in a 33-year-old male dolphin. However, Turner's young age (8–10 years) and age relative to Beethoven (13 months younger) suggest that typical mammalian age-related hearing is not a plausible explanation. Environmental noise exposure may also be ruled out, since Beethoven and Turner were housed together for nearly all of Turner's life and Beethoven's hearing appears normal. This leaves ototoxic

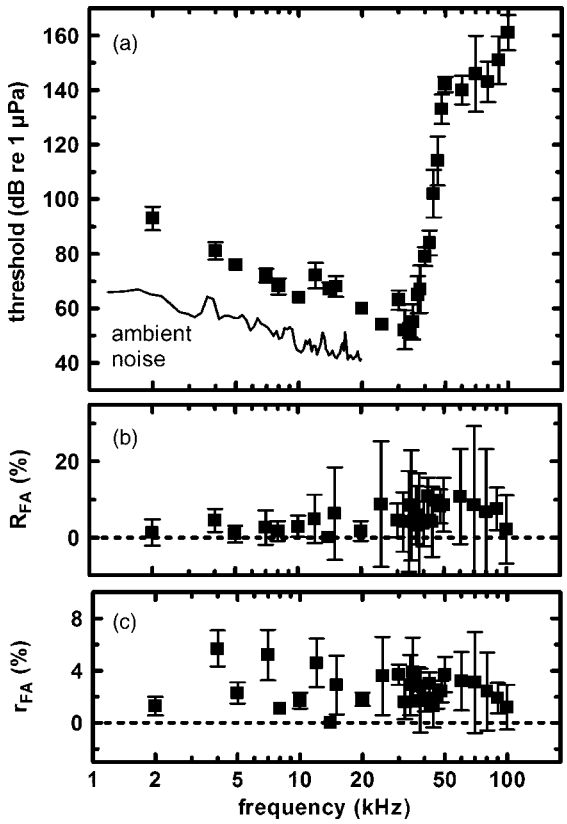


FIG. 3. Hearing thresholds, (b) R_{FA}, and (c) r_{FA} values as functions of frequency for Turner. The symbols indicate mean values, the error bars represent the 95% confidence intervals.

drug exposure as a leading candidate for the observed hearing loss.

Many drugs are toxic to the hair cells of the cochlea or vestibular system. Known ototoxic drugs include the salicylates (e.g., aspirin and aspirin-containing products), quinines, loop diuretics, aminoglycoside antibiotics (e.g., streptomycin, gentamicin, kanamycin, amikacin) and some antineoplastics (Griffin, 1988; Rybak, 1986; Tange, 1998). The extent of ototoxicity and site of damage (vestibular or cochlear) vary. For example, for the salicylates, ototoxicity occurs in approximately 1% of people receiving the drug, occurs in the cochlea, and is temporary (Jung *et al.*, 1993; Tange, 1998). Initial damage from ototoxic drugs is normally in the basal end of the cochlea, leading to high frequency hearing loss (Prosen *et al.*, 1978; Sande and Mandell, 1985; Aran *et al.*, 1995; Tange, 1998; Tan *et al.*, 2001).

To assess the possibility that exposure to ototoxic drugs was responsible for Turner's hearing loss, the health records of Beethoven and Turner were examined to determine the extent to which either had been administered ototoxic drugs. Beethoven had no history of ototoxic drug treatment; however, Turner had received aminoglycoside antibiotics. In April 1994, at six months age, Turner was diagnosed with *Nocardia* spp. infection. *Nocardia* is an infection caused by a funguslike bacterium that begins in the lungs and can spread to the brain (Turkington, 1999). Nocardial infections have a very high mortality rate (Turkington, 1999) and have been

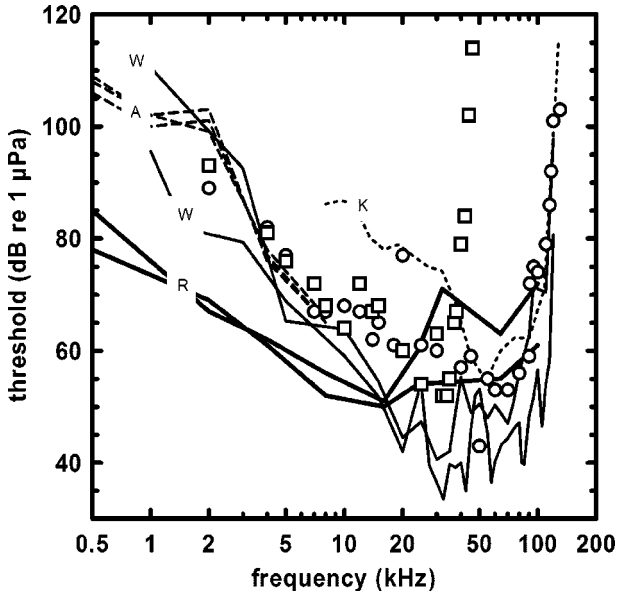


FIG. 4. Comparison between data from the present study and previously published beluga hearing thresholds. Circles—Beethoven; squares—Turner; w—White *et al.* (1978), adult male and female; A—Awbrey *et al.* (1988), adult and juvenile males and adult female; K—Klishin *et al.* (2000), adult male; and R—Ridgway *et al.* (2001), adult male and female.

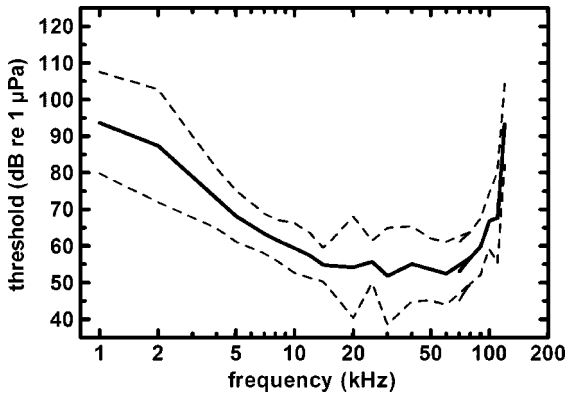


FIG. 5. Composite beluga audiogram created by collapsing the data from Fig. 4 at each frequency. Data from Klishin *et al.* (2000) and subject Turner from the present study were not included. Dotted lines represent the mean ±one standard deviation. The number of data points at each frequency were as follows: 1 kHz, six; 2–8 kHz, seven; 10–100 kHz, five; and 110–120 kHz, three.

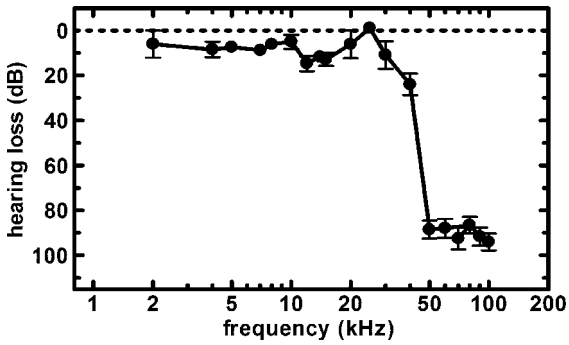


FIG. 6. Estimated hearing loss for Turner as a function of frequency. Error bars represent the 95% confidence intervals and are only shown at frequencies where data existed for both Turner and the composite audiogram shown in Fig. 5. Linear interpolation was used to estimate the ''normal'' beluga hearing thresholds at other frequencies.

reported in several cetacean species, including *Delphinapterus* (Dunn *et al.*, 2001).

Recommended treatment for *Nocardia* includes the aminoglycoside antibiotic amikacin (Dunn *et al.*, 2001), which is known to possess cochlear toxicity (Sande and Mandell, 1985; Matz, 1986). In April 1994 Turner received 8.27 mg/kg twice daily (BID) for 23 days. In October 1995, Turner received 15 mg/kg amikacin once daily (SID) for 34 days. During the course of Turner's amikacin therapy, periodic serum trough levels were documented. His 12- and 24-hour trough levels ranged between a high of 2.6 mcg/ml of serum to a low of $<$1.0 mcg/ml of serum. Peak levels of amikacin were measured during the time he received 15 mg/kg SID. The measured level was 52 mcg/ml, within the 50–60-mcg/ml range targeted for peak amikacin serum levels when utilizing SID therapy.

Amikacin, like the aminoglycoside antibiotics kanamycin and neomycin, is toxic to cochlear outer hair cells and affects those cells in the basal end of the cochlea (higher frequencies) first (Prosen *et al.*, 1978; Hawkins, 1959; Aran *et al.*, 1995; Tange, 1998). The toxicity of amikacin is similar to that of kanamycin and neomycin, with incidences of ototoxicity as high as 13% of those receiving treatment (Matz, 1986; Griffin, 1988). The relatively high amikacin dosages given to Turner and long treatment period, coupled with the observed high-frequency hearing loss, suggest that this hearing loss was a result of the amikacin treatment.

It should be pointed out that aminoglycoside antibiotics may be used to treat life-threatening infections that are resistant to other types of drugs—there may be no other choice but to use them. In the present case, Turner's severe high-frequency hearing loss must be weighed relative to the high mortality rate associated with *Nocardia*, especially in cetaceans. Recently, Bates (2003) has published data suggesting that a decrease in hearing loss induced by aminoglycosides is observed when antioxidants or iron chelator therapy is given concomitantly with aminoglycoside antibiotics. This has yet to be tested in cetaceans but might be useful to consider when infections must be treated with amikacin or other aminoglycoside antibiotics. It is interesting to note that Turner received 750 IU Vitamin E and 750 mg Vitamin C orally as a part of his routine daily dietary vitamin supplementation. It may also be noted that Turner's serum iron levels fluctuated, during the amikacin treatment periods, from his normal levels to very low levels which are common during illness in belugas.

V. CONCLUSIONS

Despite similar ages, ancestry, and environmental conditions, large differences were observed in high-frequency hearing thresholds between the two subjects. While Beethoven's thresholds were consistent with previously published data for belugas, Turner exhibited significant high-frequency hearing loss above 37 kHz, with hearing loss approaching 90 dB for frequencies above 50 kHz. An analysis of environmental factors and previous drug treatments suggests that the observed hearing loss was a result of treatment with the aminoglycoside antibiotic amikacin, which is known to be toxic to outer hair cells of the cochlea in terrestrial mammals. Amikacin and other aminoglycoside antibiotics may be used to treat life-threatening infections that are resistant to other types of drugs, so there may be no choice but to use them in certain situations. In these cases, careful dosage and/or monitoring of serum levels and possibly concomitant protective therapy may help to lower the risk of substantial hearing loss.

ACKNOWLEDGMENTS

This work supported by the U.S. Office of Naval Research Marine Mammal S&T Program and the U.S. Navy CNO(N45). We thank the staff at the Point Defiance Zoo and Aquarium for animal training and other valuable assistance.

American Speech-Language-Hearing Association (ASLHA) (**2004**). ''Types of Hearing Loss,'' http://www.asha.org/public/hearing/disorders/types.htm

Aran, J.-M., Chappert, C., Dulon, D., Erre, J.-P., and Aurousseau, C. (**1995**). ''Uptake of amikacin by hair cells of the guinea pig cochlea and vestibule and ototoxicity: Comparison with gentamicin,'' Hear. Res. **82**, 179–183.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Awbrey, F. A., Thomas, J. A., and Kastelein, R. A. (**1988**). ''Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*,'' J. Acoust. Soc. Am. **84**, 2273–2275.

Bates, D. E. (**2003**). ''Aminoglycoside ototoxicity,'' Drugs Today **39**, 277–285.

Brill, R. B., Moore, P. W. B., and Dankiewicz, L. A. (**2001**). ''Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones,'' J. Acoust. Soc. Am. **109**, 1717–1722.

Cornsweet, T. N. (**1962**). ''The staircase method in psychophysics,'' Am. J. Psychol. **75**, 485–491.

Davis, H., and Silverman, S. R. (**1978**). *Hearing and Deafness*, 4th ed. (Rinehart and Winston, New York).

Dunn, J. L., Buck, J. D., and Robeck, T. R. (**2001**). ''Bacterial diseases of cetaceans and pinnipeds,'' in *Marine Mammal Medicine*, edited by L. A. Dierauf and F. M. D. Gulland (CRC, Boca Raton, FL), pp. 309–335.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (**1961**). ''Operating characteristics, signal detectability, and the method of free response,'' J. Acoust. Soc. Am. **33**, 993–1007.

Finneran, J. J., Schlundt, C. E., Carder, D. A., Clark, J. A., Young, J. A., Gaspin, J. B., and Ridgway, S. H. (**2000**). ''Auditory and behavioral responses of bottlenose dolphins (*Tursiops truncatus*) and white whales (*Delphinapterus leucas*) to impulsive sounds resembling distant signatures of underwater explosions,'' J. Acoust. Soc. Am. **108**, 417–431.

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002a**). ''Temporary shift in masked hearing thresholds (MTTS) in odontocetes after exposure to single underwater impulses from a seismic watergun,'' J. Acoust. Soc. Am. **111**, 2929–2940.

Finneran, J. J., Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002b**). ''Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched-noise,'' J. Acoust. Soc. Am. **112**, 322–328.

Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2002c**). ''Low frequency acoustic pressure, velocity, and intensity thresholds in a bottlenose dolphin (*Tursiops truncatus*) and white whale (*Delphinapterus leucas*),'' J. Acoust. Soc. Am. **111**, 447–456.

Finneran, J. J. (**2003**). ''An integrated computer-controlled system for marine mammal auditory testing,'' SSC San Diego Technical Document 3159, June.

Finneran, J. J., Dear, R., Carder, D. A., and Ridgway, S. H. (**2003**). ''Auditory and behavioral responses of California sea lions (*Zalophus californianus*) to underwater impulses from an arc-gap transducer,'' J. Acoust. Soc. Am. **114**, 1667–1677.

Glorig, A. (**1988**). ''Damage-risk criteria for hearing,'' in *Noise and Vibration Control*, 1988 edition, edited by L. Beranek (Institute of Noise Control Engineering, Washington, DC).

Green, D. M., DeFerrari, H. A., McFadden, D., Pearse, J. S., Popper, A. N., Richardson, W. J., Ridgway, S. H., and Tyack, P. L. (**1994**). *Low-Frequency Sound and Marine Mammals: Current Knowledge and Research Needs* (National Academy, Washington, D.C.).

Griffin, J. P. (**1988**). ''Drug-induced ototoxicity,'' Br. J. Audiol. **22**, 195–210.
Hawkins, Jr., J. E. (**1959**). ''The ototoxicity of kanamycin,'' Ann. Otol. Rhinol. Laryngol. **68**, 698–714.
Johnson, C. S. (**1966**). ''Auditory thresholds of the bottlenosed porpoise (*Turisops truncatus*, Montagu),'' U.S. Naval Ordnance Test Station NOTS TP 4178.
Johnson, C. S. (**1967**). ''Sound detection thresholds in marine mammals,'' in *Marine Bio-Acoustics, Vol. 2*, edited by W. N. Tavolga (Pergamon, New York), pp. 247–260.
Johnson, C. S. (**1968**). ''Relation between absolute threshold and duration of tone pulses in the bottlenosed dolphin,'' J. Acoust. Soc. Am. **43**, 757–763.
Johnson, C. S. (**1986**). ''Dolphin audition and echolocation capacities,'' in *Dolphin Cognition and Behavior: A Comparative Approach*, edited by R. J. Schusterman, J. A. Thomas, and F. G. Wood (Erlbaum, Hillsdale, NJ).
Jung, T. T., Rhee, C. K., Lee, C. S., Park, Y. S., and Choi, D. C. (**1993**). ''Ototoxicity of salicylate, nonsteroidal anti-inflammatory drugs, and quinine,'' Otolaryngol. Clin. North Am. **26**, 791–810.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York).
Klishin, V. O., Popov, V. V., and Supin, A. Ya. (**2000**). ''Hearing capabilities of a beluga whale, *Delphinapterus leucas*,'' Aquat. Mammals **26**, 212–228.
Matz, G. J. (**1986**). ''Aminoglycoside ototoxicity,'' Am. J. Otolaryngol. **7**(2), 117–119.
Miller, H. (**1969**). ''The FROC curve: a representation of the observer's performance for the method of free response,'' J. Acoust. Soc. Am. **46**, 1473–1476.
Nachtigall, P. E. (**1986**). ''Vision, audition, and chemoreception in dolphins and other marine mammals,'' in *Dolphin Cognition and Behavior: A Comparative Approach*, edited by R. J. Schusterman, J. A. Thomas, and F. G. Wood (Erlbaum, Hillsdale, NJ).
Nachtigall, P. E., Lemonds, D. W., and Roitblat, H. L. (**2000**). ''Psychoacoustic studies of dolphin and whale hearing,'' in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 330–363.
Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing*, 2nd ed. (Academic, San Diego).
Prosen, C. A., Petersen, M. R., Moody, D. B., and Stebbins, W. C. (**1978**). ''Auditory thresholds and kanamycin-induced hearing loss in the guinea pig assessed by a positive reinforcement procedure,'' J. Acoust. Soc. Am. **63**, 559–566.
Ridgway, S. H., and Carder, D. A. (**1993**). ''High-frequency hearing loss in old (25+ years-old) male dolphins,'' J. Acoust. Soc. Am. **94**, 1830(A).
Ridgway, S. H., and Carder, D. A. (**1997**). ''Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin,'' J. Acoust. Soc. Am. **101**, 590–594.
Ridgway, S. H., Carder, D. A., Kamolnick, T., Smith, R. R., Schlundt, C. E., and Elsberry, W. R. (**2001**). ''Hearing and Whistling in the Deep Sea: Depth Influences Whistle Spectra But Does Not Attenuate Hearing by White Whales (*Delphinapterus leucas*, Odontoceti, Cetacea),'' J. Exp. Biol. **204**, 3829–3841.
Rybak, L. P. (**1986**). ''Drug ototoxicity,'' Annu. Rev. Pharmacol. Toxicol. **26**, 79–99.
Sande, M. A., and Mandell, G. L. (**1985**). ''Antimicrobial agents: the aminoglycosides,'' in *Goodman and Gilman's: The Pharmacological Basis of Therapeutics*, edited by A. Goodman-Gilman, L. S. Goodman, T. W. Rall, and F. Murad (Macmillan, New York), p. 1150.
Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). ''Temporary shift in masked hearing thresholds (MTTS) of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones,'' J. Acoust. Soc. Am. **107**, 3496–3508.
Schusterman, R. J., Barrett, R., and Moore, P. W. B. (**1975**). ''Detection of underwater signals by a California sea lion and a bottlenose porpoise: variation in the payoff matrix,'' J. Acoust. Soc. Am. **57**, 1526–1532.
Szymanski, M. D., Bain, D. E., and Kiehl, K. (**1999**). ''Killer Whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms,'' J. Acoust. Soc. Am. **106**, 1134–1141.
Tan, C.-T., Hsu, C.-J., Lee, S.-Y., Liu, S.-H., and Lin-Shiau, S.-Y. (**2001**). ''Potentiation of noise-induced hearing loss by amikacin in guinea pigs,'' Hear. Res. **161**, 72–80.
Tange, R. A. (**1998**). ''Ototoxicity,'' Adverse Drug React Toxicol. Rev. **17**, 75–89.
Turkington, C. A. (**1999**). ''Norcardiosis,'' in Gale Encyclopedia of Medicine, edited by K. Boyden and D. Olendorf (Gale Research Group, Farmington Hills, MI).
White, M. J., Jr., Norris, J. C., Ljungblad, D. K., Barton, K., and di Sciara, G. N. (**1978**). ''Auditory thresholds in two beluga whales (*Delphinapterus leucas*),'' in Hubbs/Sea World Research Institute Tech. Rept., pp. 78–109 (Hubbs Research Institute, San Diego, CA).
Yost, W. A. (**1994**). *Fundamentals of Hearing* (Academic, San Diego).

The behavioral response of mice to gaps in noise depends on its spectral components and its bandwidth

James R. Ison,[a] Paul D. Allen, Peter J. Rivoli, and Jason T. Moore
Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, New York 14627

(Received 17 January 2005; revised 10 March 2005; accepted 16 March 2005)

The purpose of these experiments was to determine whether detecting brief decrements in noise level (''gaps'') varies with the spectral content and bandwidth of noise in mice as it does in humans. The behavioral effect of gaps was quantified by their inhibiting a subsequent acoustic startle reflex. Gap durations from 1 to 29 ms were presented in five adjacent 1-octave noise bands and one 5-octave band, their range being 2 kHz to 64 kHz. Gaps ended 60 ms before the startle stimulus (experiment 1) or at startle onset (experiment 2). Asymptotic inhibition was greater for higher-frequency 1-octave bands and highest for the 5-octave band in both experiments, but time constants were related to frequency only in experiment 1. For the lowest band (2–4 kHz) neither noise decrements (experiment 1 and 2) nor increments (experiment 3) had any behavioral consequence, but this band was effective when presented as a pulse in quiet (experiment 4). The lowest frequencies in the most effective 1-octave band were one octave above the spectral region where mice have their best absolute thresholds. These effects are similar to those obtained in humans, and reveal a special contribution of wide band, high-frequency stimulation to temporal acuity. [DOI: 10.1121/1.1904387]

PACS numbers: 43.80.Lb, 43.66.Gf, 43.66.Mk [WA] Pages: 3944–3951

I. INTRODUCTION

Auditory temporal acuity in human listeners and in laboratory animals is most often assessed by measuring the detection threshold for brief quiet periods or decrements in noise level (temporal gaps) placed within a broadband noise. Average detection thresholds of about 2 to 4 ms are typically obtained in psychophysical studies in humans (e.g., Plomp, 1964), and similar gap thresholds are found in behavioral studies of gap detection in mice (Barsz *et al.*, 2002) and in other laboratory animals (chinchillas, Giraudi *et al.*, 1980; ferrets, Kelly *et al.*, 1996; gerbils, Wagner *et al.*, 2003; and rats, Ison, 1982). This uniformity in gap thresholds across mammalian species contrasts with their substantial differences in both the spectral frequencies that provide their best absolute thresholds and in the bandwidth of their effective range of hearing. The difference between species in their ability to detect particular tonal stimuli is most apparent in the comparison of humans and mice, young adult humans having their lowest absolute thresholds in the region of about 1 to 2 kHz (Brant and Fozard, 1990; Corso, 1958) and an nominal audiometric range of about 20 Hz to 20 kHz (Greenwood, 1961), while for young adult mice the best absolute thresholds are found at about 14 to 16 kHz and their nominal hearing range is about 1 kHz to 100 kHz (Fay, 1988, pp. 367–370).

The similarity of these measures of temporal acuity despite the very different spectral sensitivities of mice and human listeners is noteworthy given the reported dependence of temporal acuity in human listeners on the high-frequency spectral composition of the noise carrier and on its bandwidth (Buus and Florentine, 1985; de Filippo and Snell, 1986; Eddins *et al.*, 1992; Fitzgibbons, 1983; Fitzgibbons and Wightman, 1982; Formby and Muir, 1988; Glasberg *et al.*, 1987; Green and Forrest, 1989; Shailer and Moore, 1983, 1985; Snell *et al.*, 1994). Buus and Florentine (1985) provided the most extensive parametric description of the effect of these two variables on gap thresholds. They reported that gap sensitivity in human listeners reached a near asymptotic value for 1-octave wide bandpass noise centered at 4 kHz, that is, at the upper limit of the spectral region of high absolute sensitivity. This high level of temporal acuity was then maintained and at high levels possibly further enhanced when the center frequency increased to 8 kHz and 14 kHz, that is, into spectral regions where absolute sensitivity is relatively poor. Moreover, temporal acuity dropped precipitously as the center frequency fell to 2 kHz and below, even for a 1 kHz stimulus that is still within the spectral region of highest absolute sensitivity. They reported also that the lowest gap thresholds were obtained with a wideband noise that covered the range of all of their separate 1-octave band carriers. This latter finding was taken as suggestive evidence that the highest level of temporal acuity apparent in human listeners results from integrating less precise temporal information across separate auditory channels.

The frequencies at which gap thresholds near their asymptotic levels for bandpass noise in these studies are relatively high in the human audiogram, but these same absolute frequencies are very low in the mouse, about two octaves below the spectral region that provide their best absolute thresholds. Our goal in the present research was to determine how gap detection in the mouse is affected by the relative width and spectral characteristics of the noise band in which

[a]Corresponding author: Telephone: +1 (585) 275-8461; fax: +1 (585) 442-9216; electronic mail: jison@bcs.rochester.edu

 0001-4966/2005/117(6)/3944/8/$22.50

the gap is embedded, in comparison to the data previously obtained in human listeners.

II. METHOD

A. Subjects

The subjects were 60 young adult CBA/CaJ mice, born and raised in the vivarium at the University of Rochester but derived from breeding stock initially acquired from the Jackson Laboratory (Bar Harbor, ME). The mice were 5 to 7 months old in experiment 1 (6 male and 9 female); 3 to 5 months old in experiment 2 (4 male, 8 female); 3 to 4 months old in experiment 3 (6 female and 6 male); and 2 to 4 months old in experiment 4 (9 male and 7 female). The mice were maintained in group cages in a constant temperature and constant humidity environment, with a 12/12 light/dark cycle (lights on at 6 a.m.). Testing was conducted usually between the hours of 9 a.m. and 6 p.m. All procedures were approved by the University of Rochester Committee on Animal Resources, and were in accord with the regulations of the Public Health Service and the Federal Animal Welfare Act.

B. Apparatus

The details of the apparatus have been previously described (Ison *et al.*, 2002). The mouse was placed for testing in a wire mesh oval-shaped cage (approximately 8 cm long and 5 cm in both height and width), that was mounted on a suspended acrylic platform to which an accelerometer was attached. This assembly was placed in an anechoic chamber. The accelerometer was sensitive to the vertical force exerted by the startle reflex, and its output was integrated over a 100 ms period, beginning with the startle stimulus. These values were recorded in arbitrary voltage units that are linearly related to the downward force provided by the startle reaction. In experiment 1 the noise sources for the filtered noise carriers, for the unfiltered background masking noise, and for the unfiltered startle stimulus were independent noise generators (linear output from 100 Hz to 100 kHz) fabricated in the departmental electronic shop. In this experiment the noise bands were produced by gating the output of one noise source through a variable analog filter with 12 dB/octave attenuation, with the edges of the band being 3 dB down from the center level. In the last three experiments the source for the startle stimulus and for the bandpass noise carriers was a digital signal generator (RP2.1, Tucker-Davis Technologies, Gainesville, FL), presented unfiltered for the startle stimulus but filtered into bands for the various carriers of the gap by the RP2.1. These bands were constructed to be identical to those provided in experiment 1. The source of the masking noise was the same in all experiments. The filtered bandpass stimuli and the masking noise were presented through a high-frequency Panasonic leaf tweeter with an output that was flat from about 2 kHz to 30 kHz and declined thereafter at a rate of about 9 dB per octave. The startle stimulus was presented through a second high-frequency tweeter.

C. Procedure

In all experiments the mouse was placed into the anechoic chamber for a 5 min period and then tested with two kinds of trials: (a) baseline control trials in which the startle stimulus was presented alone, and (b) trials in which the startle stimulus immediately followed a perturbation in the acoustic background. The startle stimulus was always a 20 ms noise burst with 0.2 ms rise/fall times, presented at 110 dB SPL. To human listeners this stimulus approximates the intensity of a finger snap into the palm at a distance of 5 cm from the ear. The intertrial intervals on average were 20 s apart and were randomly chosen from a rectangular distribution that had a range from 15 to 25 s. The startle alone condition was always given twice in each block of trials while the other conditions were given just once, all presented in random order within a block. Each test day in each experiment consisted of 11 blocks of trials, the number of trials within each block depending on the numbers of conditions in each experiment.

The 1-octave noise bands were presented at the same attenuator setting that provided a level of 70 dB SPL for the 5-octave noiseband. These bandpass stimuli were always mixed with an independent wideband noise that was set at 50 dB SPL, and thus the relative level of any noise band that carried the gap information was always 20 dB above the level of the same band in the background noise. The rise/fall times of the noise at both ends of the gaps in experiments 1 and 2 and the increments in experiment 3 were near instantaneous (0.2 ms) but had a 5 ms duration in experiment 4 as in this experiment the noise bands were presented in quiet.

The same set of passbands were used in all of the experiments, namely, 2–4 kHz, 4–8 kHz, 8–16 kHz, 16–32 kHz, 32–64 kHz, and 2–64 kHz. While they were presented at a constant relative level of 20 dB SPL, these bands necessarily differed in absolute level because of their different bandwidths and because of the frequency response of the speaker. The levels of the 1-octave bands between 2 and 4 kHz, 4 and 8 kHz, 8 and 16 kHz, 16 and 32 kHz, and 32 and 64 kHz were in order, 57.2, 58.9, 61.9, 63.3, and 60.9 dB SPL. The level of the 5-octave band was 67.2 dB SPL. The geometric centers of these bands were about 2.9 kHz, 5.7 kHz, 11.4 kHz, 22.9 kHz, 45.7 kHz, and 11.4 kHz. In our colony of young adult CBA mice the mean (SD, N) ABR thresholds (dB SPL) for pure tone stimuli presented near the center of these passbands are for 3 kHz, 54.7 dB (10.1, 284); 6 kHz, 28.5 dB (8.2, 284); 12 kHz, 13.2 dB (7.3, 284); 24 kHz, 25.3 dB (11.0, 284); and 48 kHz, 33 dB (11.0, 283); and for WBN, 11.1 dB (5.2, 153).

D. Experimental designs

1. The inhibitory effects of gaps in noise bands differing in their frequency spectrum and bandwidth

This experiment was the central focus of this work, in its examination of the varied inhibitory effects of brief gaps of different durations embedded in different noise bands, with the gaps ending always 60 ms prior to the onset of the acoustic startle. Gaps placed in broadband noise provide an inhibitory gap function that rises steeply to near asymptotic levels

as the gap duration increases from 0 to about 4 ms in mice (Ison *et al.* 2002). When this method was used to test human listeners the ''threshold value for inhibition,'' defined conventionally as the duration at which inhibition was 50% of its asymptotic value, approximated the gap threshold obtained using the standard psychophysical method of constant stimuli in the same set of subjects (Ison and Pinckney, 1983).

In experiment 1 eight gap durations were presented on each of six test days (0, 2, 4, 7, 11, 16, 22, and 29 ms gaps inserted into the noise bands). One no-stimulus trial on which background activity was measured was also included in each block. A different filter setting for the noise stimuli in which the gap was embedded was used on each day, their order counterbalanced across subjects. Because gap detection improves slightly with experience in mice (Ison, 2001), prior to the beginning of this experiment the mice were given three practice days with a 10 ms gap presented at different lead times prior to the startle stimulus. In the first practice test day the noise carrier for the gap was a 70 dB wideband noise and there was no background masking noise, but for the second and third days the noise that contained the gap was either a high-frequency (16–32 kHz) or a low-frequency (4–8 kHz) band of noise mixed with the constant wideband masking noise, these counterbalanced across mice.

2. The inhibitory effects of the offset of noise bands differing in their frequency spectrum and bandwidth

This experiment was intended to determine whether the effects of the spectral composition and bandwidth on reflex inhibition observed to occur with ''complete'' gaps in experiment 1 would be found also when only the first half of a gap, that is, the noise decrement, was used as the inhibitory event. The design of experiment 2 was very similar to that of experiment 1, save that the gap was the interval between the offset of the bandpass noise and the onset of the startle stimulus, and nine gap durations were presented on each of the six test days (0, 1, 2, 4, 7, 11, 16, 22, and 29 ms). In experiment 2 a gap was placed between the offset of the noise band and the onset of the startle stimulus and its duration was simply the time interval between the two stimulus events. The time constant and the asymptotic level of the inhibitory effect of noise offset are similar to that of a complete gap of the same duration (Ison *et al.*, 1998).

3. The inhibitory effects of brief increments in the level of noise bands differing in their frequency spectrum and bandwidth

Experiment 3 was the complement of experiment 2, to determine whether the effects of spectral composition and bandwidth on asymptotic Rd′ produced by complete gaps observed in experiment 1 would be observed as well if only the ''second half'' of the gap, that is, a noise increment, was used as the inhibitory event. In experiment 3 the increments in the background noise were 20 ms in duration and they were presented 60 ms prior to the startle stimulus. This is the same lead time as the onset of noise at the end of the gap in experiment 1. Each pulse had a duration of 20 ms. The series of trials in which the seven stimulus conditions were given (including the five 1-octave noise bands, the 5-octave noise band, and a baseline condition) was repeated on each of three identical test days.

4. The inhibitory effect of brief noise bands differing in their frequency spectrum and bandwidth, when presented in quiet

The purpose of experiment 4 was to determine if each of the noise bands used in the prior experiments was audible, as would be demonstrated by their having an inhibitory effect when they were presented as pulses in a quiet background. The design of this experiment was exactly that of experiment 3, save that the stimuli were shaped by 5 ms rise and fall times and they were presented in quiet. The noise bands presented in quiet elicited startle reactions on some trials, and these trials were eliminated from subsequent analyses in order to eliminated any possible confounding between the prestimulus reaction and the reaction to the explicit startle stimuli. The criterion used to identify startle reactions to the preliminary stimuli in each mouse was that the response were more than three standard deviations above the mean background activity level for baseline control trials on which prestimuli were not presented.

E. Data analysis

The analysis of the data for reflex inhibition began with the arithmetic difference in response voltage between the mean of the two control trials and the mean for each prestimulus condition for each individual trial. (The first block of trials was not used in these analyses, in order to eliminate the sometimes unusually large responses that may appear in the first trials of a test session.) The differences were calculated within each block of trials for each subject, and were then used to calculate a standardized ''effect size'' measure (''*d*,'' Cohen, 1992) for each condition for each mouse for each test day. This measure was obtained by dividing the mean difference in ASR strength between the control condition and the prestimulus condition by the pooled standard deviation of the differences across all prestimulus conditions. This provides a d'-like response measure of the salience of each gap condition for each mouse, which we call Rd′:

$$\mathrm{Rd}' = \frac{\mathrm{ASR}_{\mathrm{Baseline\ Condition}} - \mathrm{ASR}_{\mathrm{Prestim\ Condition}}}{\sigma_{\mathrm{Pooled}}}.$$

The Rd′ measures were subjected to repeated measures analyses of variance (ANOVA) with gap duration and noise bands as within-*S* variables (using SPSS, version 12.0). The degrees of freedom for the ANOVA were adjusted for nonhomogeneity of between-cell correlations by the Huynh–Feldt procedure (Huynh and Feldt, 1976). The graphical presentation, regression analyses, and effect size calculations used GraphPad Prism software (version 4.2).

III. RESULTS

A. Experiment 1

Figure 1 depicts the mean (SEM) Rd′ values across gap duration for the six frequency bands in which the gap was

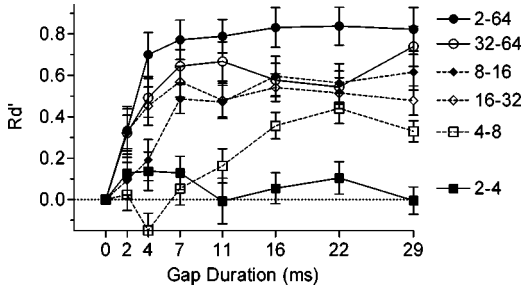


FIG. 1. Mean (SEM) levels of Rd′, showing the inhibitory effect of presenting a gap in noise just prior to the acoustic startle response in mice as a function of gap duration and the bandpass frequency of the carrier of the gaps. The gap ended 60 ms before the onset of the startle stimulus. A continuous wideband noise that was 20 dB SPL below the level of the 2–64 kHz bandpass carrier was always present (Experiment 1).

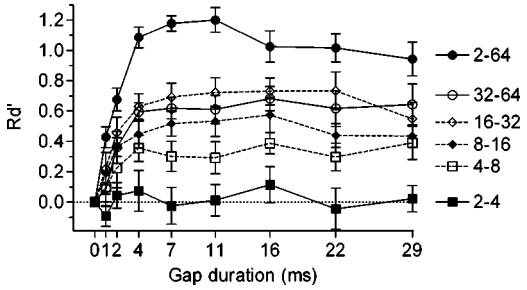


FIG. 2. Mean (SEM) levels of Rd′, showing the inhibitory effect of presenting a gap in noise just prior to the acoustic startle response in mice as a function of gap duration and the bandpass frequency of the carrier of the gaps. The gap ended with the onset of the startle stimulus. A continuous wideband noise that was 20 dB SPL below the level of the 2–64 kHz bandpass carrier was always present (Experiment 2).

inserted. The behavioral function obtained with the 5-octave bandpass noise is typical of the effect of a gap in a wideband noise for the young CBA mouse in its rapid increase to attain a near asymptotic level of reflex inhibition at about 4 ms. The 1-octave bands with upper cutoffs of 32 kHz and 64 kHz were the equal of the wideband noise in their providing an identical early onset of strong inhibition for the 2 ms gap. In contrast this 2 ms gap provided, at best, a minimal level of inhibition when it was presented in any of the three lowest 1-octave bands with their upper cutoffs at 4 kHz, 8 kHz, and 16 kHz. Even the longest gaps in the 2–4 kHz band had no behavioral effect ($p>0.2$), while the smallest gap that had a significant effect ($p<0.05$) in each of the other conditions varied according to the bandpass frequencies: 16 ms for 4–8 kHz, 7 ms for 8–16 kHz, and 2 ms for both 16–32 kHz and 32–64 kHz. The ANOVA across all 1-octave bands provided significant main effects for frequency, $F(4/56)=14.11$, $p<0.001$, $\eta_p^2=0.50$; and gap duration, $F(6/84)=12.54$, $p<0.001$, $\eta_p^2=0.47$; and a significant interaction, $F(24/336)=4.02$, $p<0.001$, $\eta_p^2=0.22$. The Rd′ means for each of these frequency bands were fit with a one-phase exponential function and the resulting rate constants, K (SE), for each frequency band are provided in Table I. This analysis indicates that the rate of approach to an asymptote was slower for the two 1-octave bands with their upper frequencies at 8 kHz and at 16 kHz. The asymptotic mean Rd′ values for the 22 and 29 ms gaps were not different between the three highest 1-octave bands ($p>0.20$), but in the adjacent bands of 8–16 kHz vs 4–8 kHz, and 4–8 kHz vs 2–4 kHz, the higher-frequency band within each pair provided a higher asymptotic performance ($p<0.02$). No 1-octave band provided as high a level of asymptotic performance as the 5-octave band of noise (the differences between the 5-octave band and each of the 1-octave bands for gaps of 22 and 29 ms were all significant, $p<0.01$). The upper limit of Rd′ is set by the average background activity, which in this experiment was 1.43 (SEM=0.08). None of the asymptotic Rd′ values approached this ceiling.

TABLE I. Time constants [K, (SE)] for the growth of Rd′ in Experiments 1 and 2.

	4–8 kHz	8–16 kHz	16–32 kHz	32–64 kHz	2–64 kHz
Exp. 1:	0.02(.04)	0.14(.04)	0.53(.19)	0.38(.12)	0.34(.08)
Exp. 2:	0.51(.14)	0.57(.12)	0.51(.10)	0.39(.07)	0.57(.11)

B. Experiment 2

Figure 2 depicts the mean (SEM) Rd′ values provided by the offset of the bandpass noise stimuli as the interval between this offset and the onset of the startle stimulus increased from 0 ms to 29 ms. As in Fig. 1, the asymptotic performance increased with the upper frequencies and the bandwidth of the noise and there was no suggestion that the offset of the lowest center frequency octave band with its upper bound set at 4 kHz had any effect on the response. The critical characteristic of these noise-offset data that contrasts with those obtained with a complete gap is that only the asymptotic levels of Rd′ were differentially affected by the different frequency bands, as in this experiment all of these behavioral functions were at their asymptotic levels for a gap duration of about 4 ms. The overall ANOVA of the 1-octave band data provided significant main effects for frequency band, $F(4/44)=10.03$, $p<0.001$, $\eta_p^2=0.48$, and for gap duration, $F(7/77)=11.89$, $p<0.001$, $\eta_p^2=0.52$, but the interaction of these two variables had a near zero effect, $F(28/308)=0.99$, $p>0.40$, $\eta_p^2=0.08$.

The first duration at which Rd′ was significant ($p<0.01$) was 1 ms for the 5-octave band; 2 ms for the three highest 1-octave bands, namely, 32–64 kHz, 16–32 kHz, and 8–16 kHz; and 4 ms for the 4–8 kHz band. The Rd′ means for each of these frequency bands were fit with a one-phase exponential function and the resulting rate constants K (SE) for each frequency band are provided in Table I. There were no systematic differences apparent in these rate constants across frequency bands within this experiment, but it should be noted that the two 1-octave bands of 4–8 kHz and 8–16 kHz had considerably shorter time constants here compared to experiment 1, while the time constants for the two higher 1-octave bands of 16–32 kHz and 32–64 kHz were similar in the two experiments. An ANOVA comparing the overall asymptotic Rd′ values (the mean values at 22 ms and 29 ms) in the 1-octave bands (ignoring the 2–4 kHz band in this analysis) provided a significant main effect for frequency band, $F(3/33)=4.79$, $p<0.01$, $\eta_p^2=0.30$, with a significant linear trend ($p<0.01$). Subsequent analyses

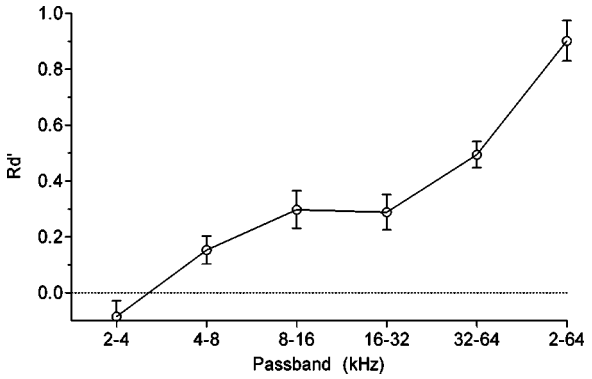


FIG. 3. Mean (SEM) levels of Rd′, showing the strength of the inhibitory effect of presenting a noise increment just prior to the acoustic startle response in mice as a function of the band-pass frequency of the pulse. The increment began 60 ms before the onset of the startle stimulus. A continuous wideband noise that was 20 dB SPL below the level of the 2–64 kHz bandpass carrier was always present (Experiment 3).

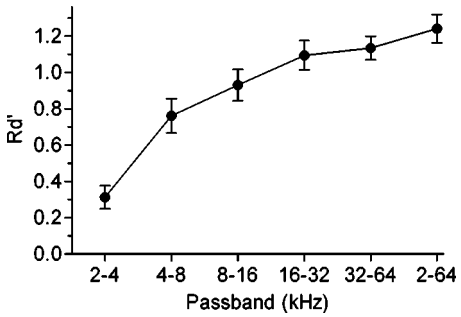


FIG. 4. Mean (SEM) levels of Rd′, showing the inhibitory effect of presenting a noise pulse just prior to the acoustic startle response in mice as a function of the bandpass frequency of the pulse. The pulse began 60 ms before the onset of the startle stimulus. These stimuli were presented in quiet (Experiment 4).

showed that the 4–8 kHz band was significantly different from both the 16–32 kHz and the 32–64 kHz bands ($p<0.02$), and marginally different from the 8–16 kHz band ($p=0.06$). The three 1-octave bands of 8–16 kHz, 16–32 kHz and 32–64 kHz did not differ from each other ($p>0.10$). The asymptotic levels for each of these 1-octave bands approximated those of experiment 1. The asymptotic level of Rd′ produced by the offset of the 5-octave band was significantly greater than for in any 1-octave band ($p<0.01$). The average Rd′ for background activity was 1.59 (SE=0.10).

C. Experiment 3

Figure 3 depicts the mean (SEM) Rd′ for each of the bandpass pulses, each presented at +20 dB over the same band within the wideband masking noise. The effect of increasing the upper frequency of each band was to produce an approximately linear increase in Rd′ that approximated the asymptotic levels obtained in the prior experiments. The ANOVA of these data provided a significant main effect for frequency band, $F(5/100)=64.73$, $p<0.001$, $\eta_p^2=0.76$, which had a significant linear component, $F(1/20)=214.88$, $p<0.001$, $\eta_p^2=0.92$. All of the pulses except the 2–4 kHz band provided significant levels of Rd′ ($p<0.01$), and the Rd′ means for all adjacent bands were significantly different from each other ($p<0.05$), save for the 16–32 kHz and 32–64 kHz bands. The average Rd′ for the no-stimulus activity condition was 1.72 (SEM=0.12).

D. Experiment 4

Figure 4 depicts the mean (SEM) Rd′ describing the effect size for inhibition for each of the bandpass pulses presented in quiet, excluding trials in which startle responses to the explicit startle stimulus were preceded by a response to the initial stimulus (see methods). The ANOVA of these data yielded a significant main effect for frequency band $F(5/55)=48.44$, $p<0.0001$, $\eta_p^2=0.82$, which had significant linear and quadratic components, $F(1/11)=105.73$, 21.35, $p<0.001$, $\eta_p^2=0.91$, 0.66. In contrast to the preceding experiments, here the lowest 1-octave band of 2–4 kHz significantly inhibited the startle response (mean Rd′=0.31, 95% confidence interval=0.17, 0.45, $p<0.001$). The lowest adjacent bands from 2–4 kHz to 4–8 kHz, to 8–16 kHz, and to 16–32 kHz were all significantly different from each other ($p<0.005$), but the higher bands of 16–32 kHz, 32–64 kHz, and 2–64 kHz did not significantly differ ($p>0.01$). The average Rd′ for the no-stimulus activity condition was 1.53 (SEM=0.09). This value was significantly higher than the Rd′ for the 2–64 kHz band, $t(11)=4.14$, $p<0.002$, indicating that inhibition was not limited by an artificial ceiling.

Gaps or noise offsets never elicit startle reactions in the mouse, but in this experiment the brief 1-octave bandpass stimuli presented in quiet did elicit a small increased level of prestartle activity while the 5-octave wideband noise provided a relatively vigorous response. The response to the 32–64 kHz 1-octave noise burst was slightly elevated compared to the no-prestimulus condition, at a mean of 452(45) compared to 302 (19), $t(11)=2.98$, $p<0.05$. In contrast, the mean response to the 5-octave 2–64 kHz stimulus (presented at 67 dB SPL) was substantial at 1414 (209), about one-third of the amplitude of the control response to the explicit startle stimulus presented at 110 dB SPL, which produced a mean response of 4165 (411). The median number of trials on which criterion startle responses occurred in response to the 2–64 kHz bandpass stimulus was 10.5 (range=0 to 25) out of a possible 30, while the median number for the 32–64 kHz stimulus was 3.0 (range 0 to 13). Few instances of a criterion response occurred for the other stimuli.

IV. DISCUSSION

Experiment 1 showed that gap thresholds and the time constants for the growth of Rd′ improved with an increase in the high-frequency components of the 1-octave widenoise band carriers extending beyond the upper end of the spectral region where the mouse is most sensitive, e.g., up to about 16 kHz; and then was at least maintained when the noise bands included the increasingly higher frequencies to which mice are increasingly less sensitive. The improvement in temporal acuity observed with higher-frequency noise bands that maintain a constant relative bandwidth is consistent with the data obtained in human listeners by Buus and Florentine (1985). In fact, the entire pattern of the beneficial effect of increasing the high-frequency content of the noise bands pre-

viously seen with human listeners was duplicated in the mouse, save for its being transposed by about two octaves. These data obtained in the mouse are also similar to those obtained in human listeners by Buus and Florentine (1985), Fitzgibbons (1983), and Snell *et al.* (1994) in their showing that increasing the relative carrier bandwidth enhances temporal acuity, as the asymptotic effect of the 5-octave band was higher than any of the 1-octave bands. These data indicate that the mouse, like the human, must also be able to combine information across frequency channels in order to achieve high levels of temporal acuity.

In experiment 2 the magnitude of Rd′ increased rapidly with an increase in delay time between the noise decrement and the onset of the startle stimulus. The asymptotic level of Rd′ was greater with an increase in the high-frequency components of the 1-octave bands of noise, and also with the increase in bandwidth from 1 to 5 octaves. But there was no change in the time constant of the increase in Rd′ with an increase in gap duration in this experiment, and near-asymptotic levels of Rd′ were reached within 4 ms for all of the carriers. If it can be assumed that noise decrement by itself captures the effect of a noise decrement at the start of a regular gap, then the data of experiment 2 suggest that the beneficial effect of the highest-frequency components of the complete gap on time constants for the growth of Rd′ results from some beneficial effect of these high frequencies for encoding the noise increment at the end of the gap, rather than differences in encoding the noise decrement at the beginning of the gap. Recent data on ''cross-frequency-channel'' gaps in the mouse (Moore *et al.*, 2005) support this hypothesis, in their showing that the combination of an onset of a relatively high-frequency noise band at the end of a gap and the offset of a relatively low-frequency band at its beginning provides better gap detection than the reverse order.

Experiment 3 was designed to isolate the effect of the noise increments at the end of the gap, by presenting each noise band as a discrete pulse at the same 60 ms lead time before the startle that followed the end of the gap in experiment 1. The inhibitory effect of these bandpass noise increments as revealed in Rd′ approximated their asymptotic effect when they were present at the end of the gap, suggesting that in this comparison gap detection and increment detection shared a common behavioral effect. This finding may be analogous in mice to those of Plomp (1964) and others, who have shown that for human listeners gap detection thresholds are fundamentally related to increment thresholds. These data suggest that the increment detection threshold for bands of noise may improve with an increase in their bandwidth in mice as well as in human listeners (Grose and Hall, 1997), and possibly in their high-frequency components as well.

Experiment 4 was undertaken to determine if all of the noise bands were audible for the mouse if presented in quiet. Although there were differences in effectiveness between the bands of noise, all provided a significant inhibitory effect. The large difference between the 2–4 kHz band and the next highest band of 4–8 kHz is possibly related to the poor sensitivity of the mouse for this very low-frequency band, but this band was at least audible. The relatively high level of startle responding to the 5-octave noise band compared to any of the 1-octave bands suggests that the sensory processes that drive the acoustic startle reflex must summate across frequency bands.

The outcome of these experiments suggests that the fundamental bases of high temporal acuity in humans and in mice are different from those that affect absolute sensitivity, and depend on the high-frequency components of the noise band in which the gap is embedded and on its relative bandwidth. While there are other hypotheses that have been persuasively proposed to explain similar findings in some earlier experiments, we suggest that these other hypotheses may be restricted in their domain to relatively low-frequency stimuli that first are not relevant for a high-frequency species like the mouse, and second, even in human listeners do not provide the high levels of temporal acuity that are routinely found for broadband stimulation that includes high frequencies.

One alternative to the ''high-frequency'' hypothesis is the ''high-absolute-bandwidth'' hypothesis, and because these two characteristics of the noise carrier are often confounded, we must concede (with Buus and Florentine, 1985, p. 166) that these hypotheses are typically difficult to distinguish even in experiments that seem to support the importance of high-frequency stimulation. In experiments performed with human listeners gaps have been embedded in noise bands as narrow as 50 Hz (e.g., Eddins *et al.*, 1992; Shailer and Moore, 1985) or as wide as 12 000 Hz (Snell *et al.*, 1994). Center frequencies of the noise band may be as low as 200 Hz (Shailer and Moore, 1983) or as high as 7 000 Hz (Buus and Florentine, 1985). The upper frequencies in the noise band may be as low as 250 Hz (Shailer and Moore, 1983) or as high as 12 000 Hz (Snell *et al.*, 1994). It is not surprising that gap thresholds obtained under these very different conditions are very different, varying from about 2 ms at best, which can be obtained with very wide noise bands that necessarily include very high frequencies (Snell *et al.*, 1994), to thresholds in excess of 30 ms at worst, obtained with narrow band noise that includes only very low frequencies (Eddins *et al.*, 1992). In one report using narrow band noise at low frequencies, an increase in the upper frequency content of noise band with the same absolute width resulted in an improved threshold (Shailer and Moore, 1985); in another report an increase in the upper frequency content of noise bands with the same absolute bandwidth resulted in no change in gap thresholds (Eddins, *et al.*, 1992); and in a third report (Snell *et al.*, 1994) an increase in the high-frequency content of a noise band with a constant absolute bandwidth resulted in an improved gap threshold with an initial increase in the upper frequency of the noise, but as the upper frequencies further increased then gap thresholds got much worse. These profound differences obtained with supposed common manipulations suggest that the gap detection for human listeners is affected by different processes as the stimulus conditions vary across the spectral range of human hearing: not all of these effects may appropriately transfer across species.

Moore (2003) described the inherent intensity fluctuations that are present in narrow band noise, and affect gap detection if the noise bandwidth is less than the bandwidth of the appropriate auditory filter. These fluctuations, produced by sparse sampling, so confuse the listener that the percep-

tual task becomes ''gap discrimination'' between long target gaps and random dips in stimulus level, rather than gap detection. In human listeners these random fluctuations would be apparent for noise bands with bandwidths below about 250 Hz and upper frequencies below about 1000 Hz, but this acoustic explanation of poor gap thresholds in ''absolute'' frequency-defined narrow band noise would not apply to mice. If the auditory filter is narrow relative to the noise band, then the acoustic fluctuations are less noticeable, and it is the time constant of the filter that limits gap detection (Moore, 2003). The time constant of a filter is proportional to the reciprocal of its absolute bandwidth, and because the bandwidth of auditory filters increases with their center frequency (Scharf, 1961), then high-frequency filters must have short time constants. Filters with long time constants ''ring'' at noise offset, and thus the programmed gap would be partially filled with continued activity. This effect must increase gap detection thresholds in humans listeners when the gap is inserted in relatively low-frequency noise. Theoretical considerations (de Boer, 1985) as well as direct measures of the rate of decay in auditory nerve firing at noise offset (Pfeiffer and Kim, 1972) indicate that auditory filter characteristics would limit gap detection only for auditory filters with their center frequencies below about 1000 Hz. Thus, this second explanation of poor gap thresholds in humans for relatively low-frequency stimuli also would not apply to gap detection in the mouse.

Noise bands with a low-pass cutoff set at 1 kHz yield gap thresholds of about 10 ms in human listeners (Fitzgibbons, 1983; Snell *et al.*, 1994) but gap thresholds continue to improve rapidly with an increase in the low-pass cutoff from 1 kHz up to about 4 kHz and continue to improve for higher frequencies, but more slowly (Fitzgibbons, 1983; Snell *et al.*, 1994). What is responsible for this further increase in sensitivity to gaps? The experiment of Eddins *et al.* (1992) indicated that the increase in absolute bandwidth can be a critical variable, but the persuasive outcome of their clever experiment appears to be restricted in applicability to the particular choice of bandwidths and upper frequencies, given the different results reported by Shailer and Moore (1985) and Snell *et al.* (1994), as described above. From a physiological perspective the enormous biological factor limiting the general applicability for the gap detection of a ''constant absolute bandwidth'' is that the position along the basilar membrane is allocated to log frequency, that is, to near constant octaves, not linear frequency (Greenwood, 1990). Adding to the apparent heavy weighting of low frequencies along the basilar membrane is the fact that the density of sensory receptors is not uniform, but is highest towards the low-frequency apex (Wright *et al.*, 1987). These characteristics of the auditory periphery must certainly be important for high absolute sensitivity, and they are probably similarly important for auditory tasks that depend on phase locking. Perhaps they are also important in affecting gap detection for low-frequency narrow band noise, but they are unlikely to play a significant role in determining the most sensitive gap detection thresholds, as these are evident only with very wide noise band signals that contain very high-frequency components.

Shailer and Moore (1983) hypothesized that the best gap detection thresholds, which they obtained for spectral frequencies above the spectral regions with the best absolute sensitivity, are the result of neural processing. The data of Zhang *et al.* (1990) support this conclusion, in their showing in chinchilla auditory nerve fibers that thresholds for gaps embedded in broadband noise improve continuously with the CF of the fiber, over a range of CF from below 2 kHz to above 8 kHz. Similarly, Eggermont (1999) found with multiunit recording in the auditory cortex of the cat that the mean thresholds for gaps in noise decreased with an increase in CF, over a range from 2 kHz up to about 20 kHz. Chinchillas and cats have a broad range of very sensitive absolute thresholds, that may extend from about 1–2 kHz to about 8–10 kHz in both species (Fay, 1988, pp. 347–352; 357–362), and so these electrophysiological data agree with the principles we propose for mice and humans, that temporal acuity improves with high-frequency stimulation and is maintained for high frequencies above those that provide the best absolute thresholds.

V. CONCLUSIONS

The behavioral findings obtained here in the mouse confirm those obtained with human listeners in showing a similar pattern of first improvement and then the maintenance of highly sensitive gap thresholds with an increase in the high-frequency components of the noise bands and with an increase in their bandwidth. Both mouse and human data, as well as the data obtained in electrophysiological experiments, show that temporal acuity is relatively independent of the mechanisms that contribute to or develop from differences in absolute sensitivity, given at least a moderate level of audibility for the carrier. Remarkably, mice and human listeners have species-specific spectral regions of peak absolute sensitivity that differ by as much as two to four octaves depending on how they are measured, but share with other mammals gap thresholds that are at least very similar, and may be identical. Differences in best frequencies for absolute thresholds vary with species-specific resonances of the ear, and match the resonances of the vocal apparatus and the resulting dominant frequencies in the vocal output in mice and in humans. In contrast, the similar values for gap detection that are found across mammalian species encourage the hypothesis that temporal acuity evolved to match the common rate of amplitude modulation that has been found in the envelope of their vocal output, and not its spectral content (Geissler and Ehret, 2002).

ACKNOWLEDGMENTS

This research was supported by USPHS NIH-NIA Research Grant No. AG09524. The authors are grateful to two anonymous reviewers for their helpful comments.

Barsz, K., Ison, J. R., Snell, K. B., and Walton, J. P. (**2002**). ''Behavioral and neural measures of auditory temporal acuity in aging humans and mice,'' Neurobiol. Aging **23**, 565–578.

Brant, L. J., and Fozard, J. L. (**1990**). ''Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging,'' J. Acoust. Soc. Am. **88**, 813–820.

Buus, S., and Florentine, M. (**1985**). ''Gap detection in normal and impaired listeners: the effect of level and frequency,'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, New York), pp. 159–179.
Cohen, J. (**1992**). ''A power primer,'' Psychol. Bull. **112**, 155–159.
Corso, J. F. (**1958**). ''Proposed laboratory standard for normal hearing,'' J. Acoust. Soc. Am. **30**, 14–23.
de Boer, E. (**1985**). ''Auditory time constants: A paradox,'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, New York), pp. 141–158.
de Filippo, C. L., and Snell, K. B. (**1986**). ''Detection of a temporal gap in low-frequency narrow-band signals by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **80**, 1354–1358.
Eddins, D. A., Hall, J. W., and Grose, J. H. (**1992**). ''The detection of temporal gaps as a function of frequency region and absolute noise bandwidth,'' J. Acoust. Soc. Am. **91**, 1069–1077.
Eggermont, J. J. (**1999**). ''Neural correlates of gap detection in three auditory cortical fields in the cat,'' J. Neurophysiol. **81**, 2570–2581.
Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL).
Fitzgibbons, P. J. (**1983**). ''Temporal gap detection in noise as a function of frequency, bandwidth, and level,'' J. Acoust. Soc. Am. **74**, 67–72.
Fitzgibbons, P. J., and Wightman, F. L. (**1982**). ''Gap detection in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **72**, 761–765.
Formby, C., and Muir, K. (**1988**). ''Modulation and gap detection for broadband and filtered noise signals,'' J. Acoust. Soc. Am. **84**, 545–550.
Geissler, D. B., and Ehret, G. (**2002**). ''Time-critical integration of formants for perception of communication calls in mice,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 9021–9025.
Giraudi, D., Salvi, R., Henderson, D., and Hamernik, R. (**1980**). ''Gap detection by the chinchilla,'' J. Acoust. Soc. Am. **68**, 802–806.
Glasberg, B. R., Moore, B. C., and Bacon, S. P. (**1987**). ''Gap detection and masking in hearing-impaired and normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1546–1556.
Green, D. M., and Forrest, T. G. (**1989**). ''Temporal gaps in noise and sinusoids,'' J. Acoust. Soc. Am. **86**, 961–970.
Greenwood, D. D. (**1961**). ''Critical bandwidth and the frequency coordinates of the basilar membrane,'' J. Acoust. Soc. Am. **33**, 1344–1356.
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Grose, J. H., and Hall, J. W. (**1997**). ''Multiband detection of energy fluctuations,'' J. Acoust. Soc. Am. **102**, 1088–1096.
Huynh, H., and Feldt, L. S. (**1976**). ''Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs,'' J. Educ. Stat. **1**, 69–82.
Ison, J. R. (**1982**). ''Temporal acuity in auditory function in the rat: reflex inhibition by brief gaps in noise,'' J. Comp. Physiol. Psychol. **96**, 945–954.
Ison, J. R. (**2001**). ''The acoustic startle response in the mouse: reflex elicitation and reflex modification by preliminary stimuli,'' *Handbook of Mouse Auditory Research: From Behavior to Molecular Biology*, edited by J. F. Willott (CRC Press, Boca Raton), pp. 59–82.
Ison, J. R., Agrawal, P., Pak, J., and Vaughn, W. J. (**1998**). ''Changes in temporal acuity with age and with hearing impairment in the mouse: A study of the acoustic startle reflex and its inhibition by brief decrements in noise level,'' J. Acoust. Soc. Am. **104**, 1696–1704.
Ison, J. R., Castro, J., Allen, P. D., Virag, T. M., and Walton, J. P. (**2002**). ''The relative detectability for mice of gaps having different ramp durations at their onset and offset boundaries,'' J. Acoust. Soc. Am. **112**, 740–747.
Ison, J. R., and Pinckney, L. A. (**1983**). ''Reflex inhibition in humans: sensitivity to brief silent periods in white noise,'' Percept. Psychophys. **34**, 84–88.
Kelly, J. B., Rooney, B. J., and Phillips, D. P. (**1996**). ''Effects of bilateral auditory cortical lesions on gap-detection thresholds in the ferret (Mustela putorius),'' Behav. Neurosci. **110**, 542–550.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing* (Academic Press, New York).
Moore, J., Allen, P., and Ison, J. (**2005**). ''Cross-spectral channel gap detection in the aging CBA mouse,'' Assoc. Res. Otolaryngol. Abstr. **28**, #432.
Pfeiffer, R. R., and Kim, D. O. (**1972**). ''Response patterns of single cochlear nerve fibers to click stimuli: descriptions for cat,'' J. Acoust. Soc. Am. **52**, 1669–1677.
Plomp, R. (**1964**). ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.
Scharf, B. (**1961**). ''Complex sounds and critical bands,'' Psychol. Bull. **58**, 205–217.
Shailer, M. J., and Moore, B. C. J. (**1983**). ''Gap detection as a function of frequency, bandwidth, and level,'' J. Acoust. Soc. Am. **74**, 467–473.
Shailer, M. J., and Moore, B. C. J. (**1985**). ''Detection of temporal gaps in band limited noise: effects of variations in bandwidth and signal-to-masker ratio,'' J. Acoust. Soc. Am. **77**, 635–639.
Snell, K. B., Ison, J. R., and Frisina, D. R. (**1994**). ''The effects of signal frequency and absolute bandwidth on gap detection in noise,'' J. Acoust. Soc. Am. **96**, 1458–1464.
Wagner, E., Klump, G. M., and Hamann, I. (**2003**). ''Gap detection in Mongolian gerbils (*Meriones unguiculatus*),'' Hear. Res. **176**, 11–16.
Wright, A., Davis, A., Bredberg, G., Ulehlova, L., and Spencer, H. (**1987**). ''Hair cell distributions in the normal human cochlea,'' Acta Oto-Laryngol., Suppl. **441**, 1–48.
Zhang, W., Salvi, R. J., and Saunders, S. S. (**1990**). ''Neural correlates of gap detection in auditory nerve fibers of the chinchilla,'' Hear. Res. **46**, 181–200.

Marine mammals and noise: Problems with root mean square sound pressure levels for transients

P. T. Madsen[a)]
Woods Hole Oceanographic Institution, Woods Hole Road 266, Woods Hole, Massachusetts 02543

(Received 10 December 2004; revised 29 March 2005; accepted 1 April 2005)

Current mitigation levels for noise transients impinging on marine mammals are specified by rms pressures. The rms measure critically relies upon choosing the size of averaging window for the squared pressures. Derivation of this window is not standardized, which can lead to 2–12 dB differences in rms sound pressure for the same wave forms. rms pressure does not represent the energy of the noise pulse and it does not prevent exposure to high peak pressures. Safety levels for transients should therefore be given by received peak–peak sound pressure and energy flux density instead of rms sound pressure levels. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1921508]

PACS numbers: 43.80.Nd, 43.80.Lb, 43.80.Ev [WA] Pages: 3952–3957

I. INTRODUCTION

The critical role of sound reception makes cetaceans susceptible to effects of manmade noise in terms of direct physiological damage, threshold shifts, masking, and disruption of normal behavior (Richardson *et al.*, 1995). The increasing concerns about the effects of underwater manmade noise on marine mammals calls for a standardized system of how to quantify and mitigate noise exposure with relevant and reproducible measures.

The magnitude of sound pressure levels in water is normally described by sound pressure on a dB scale relative to a reference rms pressure of 1 μPa (dB re 1 μPa). The nonintuitive nature of decibels, and the different reference values and properties of air and water have led to a plethora of misconceptions concerning the magnitude and potential effects of noise levels in air and water (Chapman and Ellis, 1998). An absolute dB measure should always be provided with a reference value, but it is equally important to state how the magnitude of the sound pressure was quantified. Sound pressures in underwater noise studies and bioacoustics are variously reported in terms of peak-peak, 0-peak, peak of envelope, peak-equivalent rms and rms. For the same transient wave form, levels in decibels may vary by 10 dB or more between these different measures of pressure, making comparisons futile. Thus, quantitative measures of underwater sound, and in particular, noise transients are haunted by inconsistency and lack of adequate information to reproduce and compare measurements, and there is a need for clarity and standardization (Richardson *et al.*, 1995; NRC, 2000).

The sound pressure of a continuous signal is normally parametrized by a rms measure, while the sound pressure of a transient is normally given in terms of peak pressure measures. For a pure sine wave the ratio between peak–peak and rms is 9 dB, but for aperiodic or low duty cycle signals the difference between peak–peak and rms varies widely and can often be 15 dB or more. Peak sound pressure values of transient signals are relevant measures of high level exposures with the risk of causing physical damage in auditory systems (Coles *et al.*, 1968). However, since the mammalian ear operates as an energy detector (Plomp and Bouman, 1959; Green, 1985), it also seems relevant to implement measures that include temporal integration when assessing sensation and damaging levels of transient noise.

For marine mammals, a rms level of safe exposure has been adapted in an attempt to accommodate how the animal may sense the received noise levels (NMFS, 2003). Broad band received levels of 180 dB re 1 μPa (rms) and 190 dB re 1 μPa (rms) are currently the lower limits for concern about temporary or permanent hearing impairment in cetaceans and pinnipeds (NMFS, 2003), and these levels form the basis for estimating impact radii of active sound sources at sea (e.g., Blackwell *et al.*, 2004; Tolstoy *et al.*, 2004). This paper explores the consequences of using the rms measure for safety levels of different noise transients impinging on marine mammals.

II. MATERIALS AND METHODS

Four commonly encountered transient signals (all sampled at 48 kHz) from high level underwater sources with the same modeled peak–peak received level of 189 dB re 1 μPa (pp) were chosen for analysis: (1) an on-axis version of the p1 pulse of a sperm whale usual click, (2) a 390 ms frequency modulated pulse akin to that of a ping from a mid-frequency sonar, (3) a short transient comparable to the on-axis signature from a powerful, impulse sound source such as an air gun array or an underwater explosion, and (4) the same impulse sound after propagation in a highly reverberant environment. The root of the mean of the squared pressure (rms) of a plane wave in a time window from 0 to T is given by

$$p_{\mathrm{rms}} = \sqrt{\frac{1}{T}\int_0^T p^2(t)\,dt},$$

$$\text{rms sound pressure level} = 10\log\left(\frac{1}{T}\int_0^T p^2(t)\,dt\right),$$

a)Electronic mail: pmadsen@whoi.edu

 0001-4966/2005/117(6)/3952/6/$22.50

where $p(t)$ is the instantaneous pressure (Urick, 1983). The analysis window is critical for rms measures of transient signals, the longer it is the lower the rms value will be.

In research on auditory traumas in humans, impulse wave forms are often modeled by a Friedlander wave that describes the idealized signature of a zero-rise time impulse (Hamernik and Hsueh, 1991). The rms value of such an impulse can be computed using different temporal definitions that relates to nulls or amplitude thresholds in the wave form (For review see, Hamernik and Hsueh, 1991). However, underwater noise pulses seldom render themselves suited for the temporal measures derived for the Friedlander wave. The *D* duration, which is given by the −10 dB end points relative to the peak of the envelope of the wave form, has been applied to determine the durations of biological transients (Møhl *et al.*, 1990). The envelope is computed by taking the absolute value of the analytical signal (Hilbert transformed wave form, relating the real and imaginary parts of the analytical signal) (Randall, 1987). As a variation of this approach, Madsen *et al.* (2002) and Møhl *et al.* (2003) used −3 dB end points relative to the peak of the envelope when computing rms measures of reverberant air gun pulses and p1 pulses in sperm whale clicks.

For signals with a good signal to noise ratio (SNR), a more common approach is to determine the duration of transients by using the relative energy in a window that incorporates the entire signal wave form along with short samples of noise on either side. In this approach the duration is often given by the part of the window that makes up 90% of the total cumulative energy in the window including the sound pulse (Malme *et al.*, 1986; Blackwell *et al.*, 2004). For short duration, well-defined clicks from toothed whales a 97% energy approach has also been implemented (Au, 1993; Madsen *et al.*, 2004). To test the effects of these temporal definitions on the duration of different transients, the −3 dB, −10 dB, 90%, and 97% approaches have been applied to the four transients signal types.

Acoustic impact is not only given by peak pressure, but also by the energy flux density of the sound pulse (Ward, 1997). The energy flux density or the sound exposure level of a sound pulse propagating as a plane wave in an unbounded medium is given by the time integral of the pressure squared (Urick, 1983; McCauley *et al.*, 2003). The energy flux density in dB re 1 μPa^2 s$^{\mathrm{i}}$ of transients can thus be approximated by 10 log to the time integral of the squared pressure over the duration of the pulse (Young, 1970), which for the same duration, *T*, is simply the rms level (in dB)+10 log(*T*):

$$\text{Energy flux density} = 10\log\int_0^T p^2(t)\,dt = 10\log\left(\frac{1}{T}\int_0^T p^2(t)\,dt\right) + 10\log(T),$$

where *T* is the window length in seconds. This estimation of energy flux density is in line with Finneran *et al.* (2002b) based on the assumption of individual pressure measurements of a plane wave. The intensity of a sound field is given by the product of the pressure and the particle velocity components divided by the specific acoustic impedance of the medium (Urick, 1983). But since marine mammals only seem to detect the pressure component of the sound field (Kastak and Schusterman, 1998; Hastings, 2004), the above-given formula can be used when assuming exposure to a plane wave well in the far-field of the sound source (Finneran *et al.*, 2002a). This approach has accordingly been used to compare the energy flux density of the four pulse types with identical peak–peak pressure, but with varying durations and rms levels.

III. RESULTS AND DISCUSSION

The p1 wave form of an on-axis sperm whale usual click is shown in Fig. 1(a), with its envelope in Fig. 1(b) and the cumulative energy flux density in Fig. 1(c). Figures 2(a)–2(c) is for a 390 ms sonar pulse and Fig. 3(a) depicts a custom generated single cycle 189 dB (pp) transient representative of the signature of an impulse sound source such as an air gun or a chemical explosive. Figure 3(b) displays the wave form of the same impulse with similar peak–peak received level as in Fig. 3(a), but in this case the pulse has propagated in a highly reverberant environment.

Figures 1(b) and 1(c) illustrate that the duration of the sperm whale p1 pulse can vary between 47 and 125 μs depending on how the duration is derived. The duration derived from −3 dB re peak of the envelope covers less than a full cycle of the wave form, so it is not surprising that this duration measure renders the highest rms level of 183 dB re 1 μPa (rms). All the three other duration measures are approximately twice as long and render essentially identical rms received levels 2–3 dB lower. However, the energy flux density of the pulse is within 1 dB around 141 dB re 1 μPa^2 s for the four duration measures. Consequently, for short, well-defined transients such as odontocete clicks with good SNR, the rms measure is quite robust and not very sensitive to the criterion used to establish the integration window, except that the −3 dB measure seems to lead to a rms sound pressure level that is significantly higher (2–3 dB) than the three others (Table I).

When comparing the different duration measures for a very different manmade transient like the sonar ping in Fig. 2(a), it is seen that the −3 dB approach again yields a rms level that is 2 dB higher than the three others, which relates to the higher average squared pressure in the −3 dB window than in a larger window where the sound pressures fluctuates more due to interference of multipaths [Fig. 2(b)]. As is the case for the sperm whale click, the −3 dB duration covers such a small part of the actual wave form that its use cannot be justified [Fig. 2(a)]. This argument is strengthened by the fact that the energy flux when using the −3 dB measure is 25 times smaller (14 dB) than when using the three other duration measures including a much larger fraction of the pulse (Table I).

Although the sonar pulse of Fig. 2(a), has the same peak–peak received level as the sperm whale click, its rms sound pressure level is between 177 and 179 dB depending on the duration used (the variance is caused by interference of multipaths). Thus, the sperm whale p1 pulse exceeds the safety limit of 180 dB re 1 μPa (rms), but the sonar ping with the same peak–peak pressure level does not. The

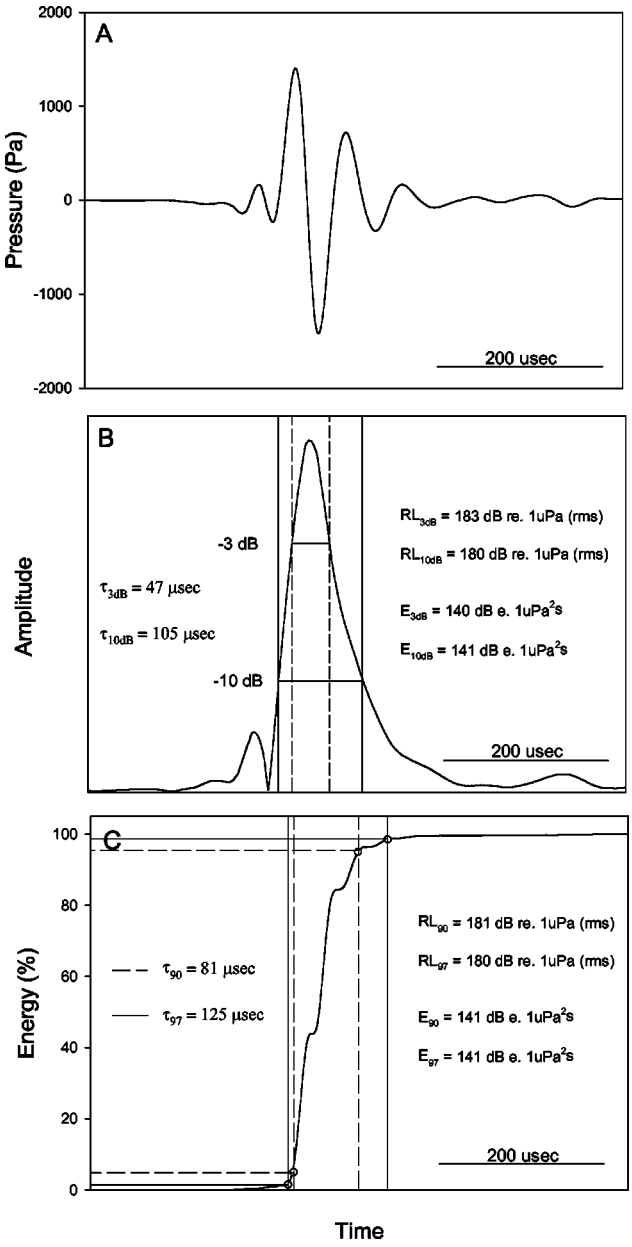


FIG. 1. (A) Wave form of p1 pulse of sperm whale click with a received sound pressure level of 189 dB re 1 μPa (pp). (B) Envelope of the wave form shown in (a). −3 and −10 dB levels for the durations of $t_{-3\ \mathrm{dB}}$ and t_{-10} dB are shown by dotted and solid gray lines. The resulting rms [dB re 1 μPa (rms)] and energy flux density (dB re 1 μPa^2 s) levels are provided for each of the duration measures. (C) The relative cumulative energy of the wave form in Fig. 1(a). The duration measures t_{90} and t_{97} is given by the windows containing 90% and 97% of the total relative energy in a window including the sound pulse.

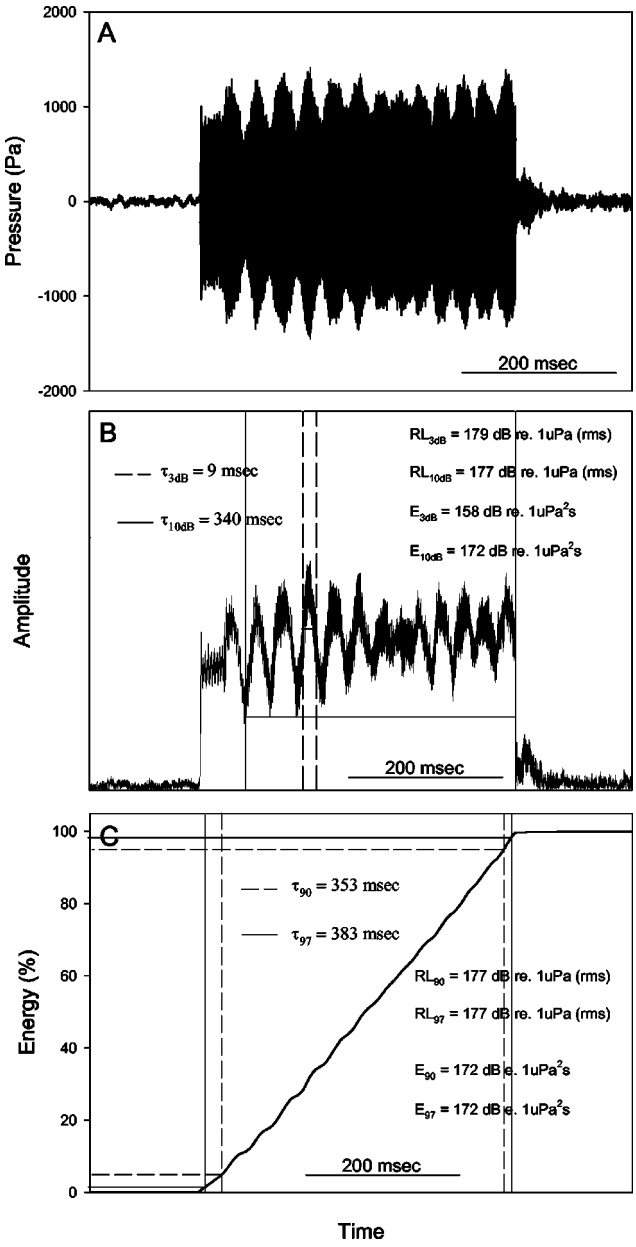


FIG. 2. (A) Wave form of $p1$ pulse of a mid-frequency sonar pulse with a received sound pressure level of 189 dB re 1 μPa (pp). (B) Envelope of the wave form shown in Fig. 2(a). −3 and −10 dB levels for the durations of $t_{-3\ \mathrm{dB}}$ and t_{-10} dB are shown by dotted and solid gray lines. The resulting rms (dB re 1 μPa (rms) and energy flux density (dB re 1 μPa^2 s) levels are provided for each of the duration measures. (C) The relative cumulative energy of the wave form in Fig. 1(a). The duration measures t_{90} and t_{97} are given by the windows containing 90% and 97% of the total relative energy a window including the sound pulse.

multipath-induced pressure fluctuations of the sonar ping lead to a lower average squared pressure than the effectively single cycled sperm whale pulse. However, when looking at the energy flux density of the sonar ping, using duration measures that essentially cover the pulse, the sonar ping is seen to carry 1000 times (30 dB) more energy than the sperm whale click for the same peak–peak received sound pressure. Thus, if the peak–peak pressure received levels of the two transients were considered, they would have an equal impact on the exposed animal. If the rms measures are used, no matter how the averaging duration is determined, the sperm whale click will exceed the 180 dB re 1 μPa (rms) limit, while the sonar ping will not, despite the fact that it is carrying more energy than the sperm whale click by three orders of magnitude (Table I). It is therefore not reasonable to compare the acoustic impact of a mid-frequency sonar pulse with that of a sperm whale click (Møhl, 2002).

The impulse sound in Fig. 3(a) has almost the same duration as the sperm whale click and about the same rms sound pressure levels of around 182 dB. Accordingly, this impulse with a similar peak–peak received sound pressure level as the three other transients of 189 dB re 1 μPa (pp),

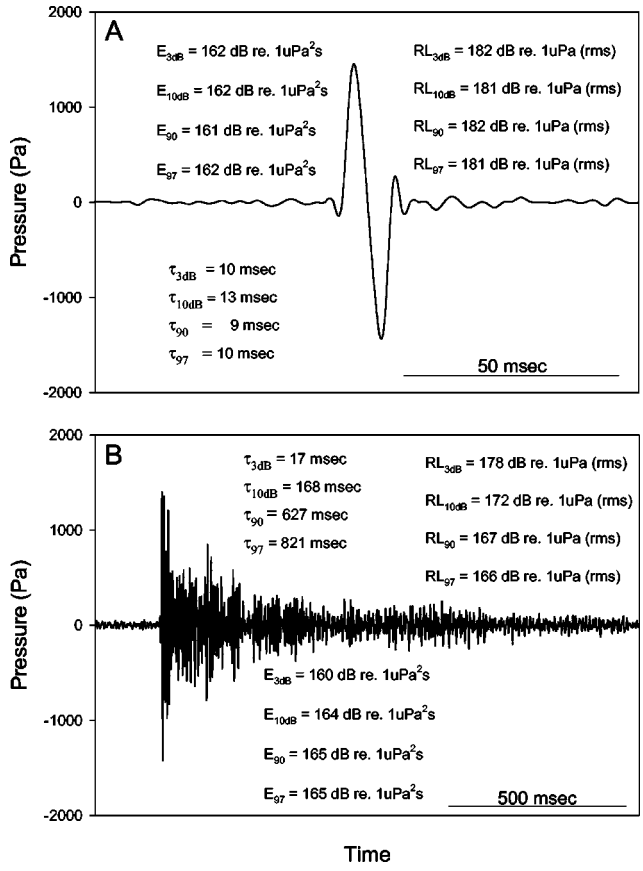


FIG. 3. (A) Transient mimicking the far-field version of a sound pulse produced by an impulse sound source in an acoustic free field. Duration, rms, and energy measures are calculated by the same means as in Figs. 1 and 2. (B) Slowly decaying transient mimicking the situation where the impulse of (a) has been propagating in a highly reverberant environment. Duration, rms, and energy measures are calculated by the same means as in Figs. 1, 2, and 3(a).

exceeds the 180 dB re 1 μPa (rms) limit. However, its energy flux density is 100 times larger than for the sperm whale click of identical peak–peak and rms received levels [Fig. 3(a)]. Thus, if energy flux density is not taken into account, the impulse would be regarded as having the same acoustical impact on an animal as the sperm whale pulse, and a larger impact than the sonar ping having 10 times (10 dB) higher energy flux density (Table I).

The wave form of Fig. 3(a) mimics the situation of a pressure wave propagating from an impulsive sound source in deep water, approaching the situation for an acoustic free-field. When such impulse sounds propagate in a highly reverberant environment such as shallow water, the energy becomes spread in time due to the variety of path lengths and group velocities supported (Greene and Richardson, 1988) as depicted in Fig. 3(b). In this case the −3 dB criterion window clearly does not cover the full extent of the pulse and neither does the −10 dB window. The choice of integration window of this slowly decaying pulse greatly affects the rms measures. However, even the highest rms measure of 178 dB re 1 μPa (rms) for this pulse, achieved with a window derived by the −3 dB criterion, is lower than the 180 dB re 1 μPa (rms) limit. Hence, multipath propagation plays an important role in determining whether the rms level received at the animal is considered too high or not, even if the energy is invariant. When using the rms measure for a transient noise pulse like the one displayed in Fig. 3(b), it is evident that the method of deriving the window may result in rms sound pressure levels that vary by as much as 12 dB (Table I). If the 90% energy measure is used for the displayed pulse, giving a window length of 627 ms, a pulse with a received peak–peak level of 202 dB re 1 μPa (pp) would still not exceed the limit of 180 dB re 1 μPa (rms). Consequently, long, fixed averaging times for calculation of rms sound pressures can yield very short safety radii around a noise source. Unless there is a specified protocol for determining the duration, it is possible to manipulate the rms level by varying the averaging window: the longer the averaging time, the lower the rms level. Measures for mitigation of sound exposure should not leave room for such analytical freedom.

The energy flux density measures the energy flow per unit area received by the animal. With the signal of Fig. 3(b), the animal is actually exposed to twice as much sound energy (3 dB) as compared to exposure to the pulse of Fig. 3(a). If the peak–peak sound pressure level is considered to indicate exposure, the pulses of Figs. 3(a) and 3(b) would be considered to have the same impact. If the rms measure was used, the pulse in Fig. 3(a) would exceed the 180 dB re 1 μPa (rms) limit, whereas the pulse of Fig. 3(b) would not even though the animal is exposed to two times the acoustic energy by the pulse in Fig. 3(b). It is also apparent that for the energy measure, durations that cover as much of the pulse as possible given the signal to noise ratio provide the highest number, which is the opposite of the rms measure.

Energy flux density is therefore a better measure for safe exposure levels than rms measures as the energy unit takes into account the overall acoustic energy impinging on the animal per unit area (McCauley *et al.*, 2003). Ears of terrestrial mammals generally integrate sound intensity over a time window of some 200 ms (Plomp and Bouman, 1959; Green, 1985), and the same appears to be the case for cetaceans at low frequencies (Johnson, 1968). It seems therefore reasonable to use 200 ms as the maximum integration time from a detector or sensation point of view (Madsen *et al.*, 2002).

TABLE I. Pulse numbers refers to the pulses displayed in Figs. 1 to 3(b). RL_{pp} is the received peak–peak sound pressure in dB re 1 μPa (pp). t provides the different duration measures in ms. rms provides the root-mean-square sound pressure in dB re 1 μPa (rms) for each of the duration measures. E gives the energy flux density in dB re 1 μPa^2 s for each the duration measures.

Pulse	RL_{pp}	$t_{3\ dB}$	$t_{10\ dB}$	t_{90}	t_{97}	$rms_{3\ dB}$	$rms_{10\ dB}$	rms_{90}	rms_{97}	$E_{3\ dB}$	$E_{10\ dB}$	E_{90}	E_{97}
1	189	0.047	0.105	0.081	0.125	183	180	181	180	140	141	141	141
2	189	9	340	353	383	179	177	177	177	158	172	172	172
3A	189	10	13	9	10	182	181	182	181	162	162	161	162
3B	189	17	168	627	821	178	172	167	166	160	164	165	165

This will lead to a 3 and 0.5 dB reduction for the pulses of Figs. 2 and 3(b), respectively. However, in terms of hearing impairment due to a single, high level impulse, it has been established that the safety threshold for humans scales as $10\log(T)$, where T is the exposure duration, even if T is much longer than 200 ms (Ward, 1997). Since this issue remains to be clarified for marine mammals, it may seem reasonable to apply a conservative approach and provide energy flux density integrated both over the entire pulse duration and with a 200 ms integration time if the actual duration is longer than that. Such measures should additionally be accompanied by a figure of the wave form, and information about the recording bandwidth and the duration used for integrating the pressure squared [as stipulated by the ANSI standard for noise exposure (ANSI, 1994)].

Impulses can have very high peak sound pressure levels, but carry very little energy (Price and Wansack, 1989). Since physical damage and impairment of the auditory system is caused both by high peak pressure and energy flux (Ahroon *et al.*, 1996; Finneran *et al.*, 2002a, b; Ward, 1997), safety limits for sound exposure should include both a maximum received energy flux level along with a maximum received peak–peak pressure level. Such a protocol addresses concerns for physical damage due to short high pressure pulses as well as the effects of longer, high-energy transients with lower peak pressures.

It is concluded that rms safety measures are unsuited as a stand-alone mitigative measure for transient noise effects on marine mammals irrespective of what the absolute level is [currently 180 dB re 1 μPa (rms) for cetaceans]. In line with Finneran *et al.* (2002a, b), it is recommended that levels set to mitigate sound exposure of marine mamals include a maximum peak–peak received sound pressure level in concert with a maximum received energy flux level (McCauley *et al.*, 2003). It is suggested that the energy flux is calculated by using the 90% energy approach for derivation of the duration (Malme *et al.*, 1986; Blackwell *et al.*, 2004), since the 97% criterion requires high signal to noise ratios, and the -3 and -10 envelope criteria underestimate the durations of slowly decaying transients. It is beyond the scope of this paper to discuss the absolute levels for mitigation of received peak–peak pressure and energy flux density, but there is an urgent need for a careful assessment of such in light of anatomical, physiological, and behavioral data for different marine mammal species.

ACKNOWLEDGMENTS

W. W. L. Au, D. Cato, C. Greene, M. Johnson, A. Michelsen, B. K. Nielsen, P. Tyack, M. Wahlberg, and W. M. X. Zimmer provided constructive critique on previous versions of the manuscript. Work supported by Contract No. 1435-01-02-CA-85186 from the US Minerals Management Service, Grant No. DACA72-01C-0011 from SERDP, and Grant No. SAIC440000 from SAIC. This publication is Contribution No. 11331 from the Woods Hole Oceanographic Institution.

[i]For a plane wave in an unbounded medium, the energy flux density in dB re 1 μPa^2 s can be converted to J/m^2 by dividing the summed squared pressure on a linear scale by the specific impedance Z (sound speed$\times$density) of the medium. For example, 182 dB re 1 μPa^2 s$=(1257\ \mathrm{Pa_{rms}})^2$ s/(1500 m/s $\times$1040 kg/m^3)$=$0 dB re 1 J/m$^2=$1 J/m^2.

Ahroon, W. A., Hamernik, R. P., and Sheau-Fang, L. (**1996**). ''The effects of reverberant blast waves on the auditory system,'' J. Acoust. Soc. Am. **100**, 2247–2257.

ANSI (**1994**). *American National Standard Acoustical Terminology, 1994*, ANSI S1.1-1994 (Acoustical Society of America, New York), p. 9.

Au, W. W. L. (**1993**). The Sonar of Dolphins (Springer, New York).

Blackwell, S. B., Lawson, J. W., and Williams, J. T. (**2004**). ''Tolerance by ringed seals (*Phoca hispida*) to impact pipe-driving and construction sounds at an oil production island),'' J. Acoust. Soc. Am. **115**, 2346–2357.

Chapman, D. M. F., and Ellis, D. D. (**1998**). ''The elusive decibel: Thoughts on sonars and marine mammals,'' Can. Acoust. **26**, 29–31.

Coles, R. R. A., Garinther, G. R., Hodge, D. C., and Rice, C. G. (**1968**). ''Hazardous exposure to impulse noise,'' J. Acoust. Soc. Am. **43**, 336–343.

Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2002a**). ''Low-frequency acoustic pressure, velocity, and intensity thresholds in a bottlenose dolphin (Tursiops truncatus) and white whale (Delphinapterus leucas),'' J. Acoust. Soc. Am. **111**, 447–456.

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002b**). ''Temporary shift in masked hearing thresholds in odontocetes after exposure to single underwater impulses from a seismic watergun,'' J. Acoust. Soc. Am. **111**, 2929–2940.

Green, D. M. (**1985**). ''Temporal factors in psychoacoustics,'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer, New York), pp. 120–140.

Greene, Jr., C. R., and Richardson, W. J. (**1988**). ''Characteristics of marine seismic survey sounds in the Beaufort Sea,'' J. Acoust. Soc. Am. **83**, 2246–2254.

Hamernik, R. P., and Hsueh, K. D. (**1991**). ''Impulse noise: Some definitions, physical acoustics and other considerations,'' J. Acoust. Soc. Am. **90**, 189–196.

Hastings, M. (**2004**). ''Noise exposure metrics for auditory and non-auditory damage in Aquatic animals,'' J. Acoust. Soc. Am. **116**, 2533.

Johnson, C. S. (**1968**). ''Relation between absolute threshold and duration-of-tone pulses in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **43**, 757–763.

Kastak, D., and Schusterman, R. J. (**1998**). ''Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology,'' J. Acoust. Soc. Am. **103**, 2216–2228.

Madsen, P. T., Kerr, I., and Payne, R. (**2004**). ''Echolocation clicks of two free-ranging delphinids with different food preferences: False killer whales (Pseudorca crassidens) and Risso's dolphin (Grampus griseus),'' J. Exp. Biol. **207**, 1811–1823.

Madsen, P. T., Møhl, B., Nielsen, B. K., and Wahlberg, M. (**2002**). ''Male sperm whale behavior during exposures to distant seismic survey pulses,'' Aquatic Mammals **28**, 231–240.

Malme, C. I., Smith, P. W., and Miles, P. R. (**1986**). ''Characterisation of geophysical acoustic survey sounds,'' OCS Study. Prepared by BBN Laboratories Inc., Cambridge, for Battelle Memorial Institute to the Department of the Interior-Mineral Management Service, Pacific Outer Continental Shelf Region, Los Angeles, CA.

McCauley, R. D., Fewtrell, J., Duncan, A. J., Jenner, C., Jenner, M. J., Penrose, J. T., Prince, R. I. T., Adhitya, A., Murdoch, J., and McCabe, K. (**2003**). ''Marine seismic surveys: Analysis and propagation of air-gun signals; and effects of exposure on humpback whales, sea turtles, fishes and squid. Environmental implications of offshore oil and gas development in Australia: Further research,'' Australian Petroleum Production Exploration, Association, Canberra, pp. 364–521.

Møhl, B. (**2002**). ''Sperm whale rivals tactical sonar with source levels at 235 dB,'' European Cetacean Society Newsletters **42**, 41–42.

Møhl, B., Surlykke, A., and Miller, L. A. (**1990**). ''High intensity narwhal click,'' in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 295–304.

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). ''The monopulsed nature of sperm whale clicks,'' J. Acoust. Soc. Am. **114**, 1143–1154.

NMFS (**2003**). ''Taking marine mammals incidental to conducting oil and gas exploration activities in the Gulf of Mexico,'' Federal register 68, 9991–9996.

NRC (**2000**). *Marine Mammals and Low-frequency Sound* (National Academy Press, Washington).

Plomp, P., and Bouman, M. A. (**1959**). ''Relation between hearing theshold and duration for tone pulses,'' J. Acoust. Soc. Am. **31**, 749–758.
Price, G. R., and Wansack, S. (**1989**). ''Hazard from an intense midrange impulse,'' J. Acoust. Soc. Am. **86**, 2185–2191.
Randall, R. B. (**1987**). *Frequency Analysis* (Brüel and Kjær, Nærum, Denmark), 344 pp.
Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, London).
Tolstoy, M., Diebold, J. B., Webb, S. C., Bohnenstiehl, D. R., Chapp, E., Holmes, R. C., and Rawson, M. (**2004**). ''Broadband calibration of R/V Ewing seismic sources,'' Geophys. Res. Lett. **31**, L14310.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula, Los Altos).
Ward, W. D. (**1997**). ''Effects of high intensity sound,'' *Encyclopedia of Acoustics* (Wiley, New York), pp. 1497–1507.
Young, R. W. (**1970**). ''On the energy transported with a sound pulse,'' J. Acoust. Soc. Am. **47**, 441–442.

Effects of exposure to seismic airgun use on hearing of three fish species

Arthur N. Popper[a)]
Department of Biology, Neuroscience and Cognitive Science Program, and Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742

Michael E. Smith
Department of Biology and Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742

Peter A. Cott
Fisheries and Oceans Canada, Yellowknife, NT, Canada X1A 1E2

Bruce W. Hanna
Fisheries and Oceans Canada, Yellowknife, NT, Canada X1A 1E2

Alexander O. MacGillivray
JASCO Research Ltd., Victoria, BC, Canada V8Z 7X8

Melanie E. Austin
JASCO Research Ltd., Victoria, BC, Canada V8Z 7X8

David A. Mann
College of Marine Science, University of South Florida, St. Petersburg, Florida 33701

(Received 7 December 2004; revised 28 February 2005; accepted 16 March 2005)

Seismic airguns produce considerable amounts of acoustic energy that have the potential to affect marine life. This study investigates the effects of exposure to a 730 in.^{3} airgun array on hearing of three fish species in the Mackenzie River Delta, the northern pike (*Esox lucius*), broad whitefish (*Coregonus nasus*), and lake chub (*Couesius plumbeus*). Fish were placed in cages in the 1.9 m of water and exposed to five or 20 airgun shots, while controls were placed in the same cage but without airgun exposure. Hearing in both exposed and control fish were then tested using the auditory brainstem response (ABR). Threshold shifts were found for exposed fish as compared to controls in the northern pike and lake chub, with recovery within 24 hours of exposure, while there was no threshold shift in the broad whitefish. It is concluded that these three species are not likely to be substantially impacted by exposure to an airgun array used in a river seismic survey. Care must be taken, however, in extrapolation to other species and to fishes exposed to airguns in deeper water or where the animals are exposed to a larger number of airgun shots over a longer period of time. [DOI: 10.1121/1.1904386]

PACS numbers: 43.80.Nd, 43.80.Lb, 43.64.Wn, 43.50.Qp, 43.30.Ma [WA] Pages: 3958–3971

INTRODUCTION

There is growing interest in the effects of anthropogenic (human-generated) sounds on marine mammals (e.g., Myrberg, 1980; Richardson *et al.*, 1995; NRC, 2003, 2005; Wartzog *et al.*, 2004) and fishes (Popper, 2003; Popper *et al.*, 2004). The continuum of potential effects on such animals could range from immediate death to no response whatsoever. In between are a range of effects that may include damage to various body tissues that could impair or ultimately kill the animal, temporary or permanent hearing threshold shift, changes in behavior because animals try to avoid the sound, and behavioral effects resulting from an animal not being able to hear biologically important environmental sounds or communication sounds from conspecifics (e.g., Richardson *et al.*, 1995; Popper, 2003; Popper *et al.*, 2004; Wartzog *et al.*, 2004).

A wide range of anthropogenic sound sources are present in the marine and freshwater environments including shipping, high power sonar, and echo sounders. Ambient levels resulting from such sounds appear to be increasing. However, it is very difficult to quantify this increase due to the diversity of sound sources in the marine environment and the very limited data available on underwater sounds in most parts of the world (NRC, 2003; Wartzog *et al.*, 2004).

Airguns are widely used for marine-based seismic exploration by the oil and gas industry. These devices produce a compressed air bubble that collapses under the pressure of water causing a sharp concussive ''explosion.'' The peak sound levels of individual airguns are as high as 230 dB (re 1μPa) at a range of 1 meter from the source. Arrays of airguns are trailed behind a vessel and put out frequent ''shots.'' The sounds reflect off geologic formations below the water

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

bottom and are detected by long streams of hydrophones towed behind the vessel. By measuring time of arrival and other characteristics of the reflected signal, it is possible to predict the presence of oil or gas in the sea bottom.

Despite the increasing interest of scientists, regulators, and environmental groups in anthropogenic sounds, there are very few experimental data that directly address how these sources affect animals. Data for fishes show that exposure to moderately loud noises can result in temporary hearing loss (called Temporary Threshold Shift, TTS) in a few species that have been studied including goldfish (*Carassius auratus*) and other fishes specialized for hearing (Popper and Clarke, 1976; Scholik and Yan, 2001, 2002; Amoser and Ladich, 2003; Amoser *et al.*, 2004; Smith *et al.*, 2004a, 2004b) (see Popper and Carlson, 1998 and Popper *et al.*, 2003 for reviews of fish hearing). Three studies using higher intensity sounds have also shown damage to the sensory hair cells of the inner ear, the cells responsible for transducing sound into neural impulses (Enger, 1981; Hastings *et al.*, 1996; McCauley *et al.*, 2003).

In the only published study to examine the direct effects of an airgun on fish physiology, McCauley *et al.* (2003) determined the effects of exposure to an airgun on the structure of fish ears. They found that exposure to multiple shots from an airgun over several hours produced damage to the sensory epithelia of the saccule, the major auditory end organ of the ear, in a group of caged pink snapper (*Pagrus auratus*), an Australian marine fish. Evidence for damage showed up as early as 18 hours post-exposure and was very extensive when fish were examined 58 days post-exposure as compared to controls.

McCauley *et al.* (2003) did not, however, test whether there were any effects on fish hearing. Indeed, the effect of anthropogenic sources on hearing is an important question since it is possible to have TTS without any permanent effects on the structure of the ear. Such TTS has the potential to put a fish in danger, since it may not hear the sounds of predators, mates, or the environment.

In the current study we examined the effects of exposure to a seismic airgun array on three species of fish found in the Mackenzie River Delta near Inuvik, Northwest Territories, Canada, an area in which there have been extensive land-based and marine seismic surveys to map rich gas and oil reserves. Considerable concern has arisen among the local population and regulators regarding proposed riverine seismic surveys and whether the sounds from the airguns could negatively impact fisheries resources (Cott *et al.*, 2003). The three species chosen for study not only represent the diversity of the fish fauna of the Mackenzie Delta region, but they also represent diversity in hearing structures found among fishes.

The species included a hearing specialist, the lake chub (*Couesius plumbeus*), and two fishes without known hearing specializations, the northern pike (*Esox lucius*), and a coregonid, the broad whitefish (*Coregonus nasus*). Broad whitefish were selected because they are one of the most important species to the aboriginal subsistence fisheries in the lower Mackenzie River and Delta (Tallman and Reist, 1997).

Fish were exposed to shots from a 730 $in.^3$ airgun array using a paradigm that would produce an exposure comparable to a worst case scenario that a fish would experience from a seismic survey in a river. The fish were tested post-exposure to measure TTS in comparison to control fish, and whether there was recovery from any TTS that was found.

METHODS

Work was done at the Mackenzie River Delta using the Fisheries and Oceans Canada facilities in Inuvik, NT. Lake chub and young of the year (YOY) northern pike were collected using beach seines along the river bank in water depths up to 1.5 m, while broad whitefish and adult northern pike were captured with short set monofilament gill nets (3.81 cm [1.5 in.]–13.97 cm [5.5 in.]) in 1.0–3.0 m of water. The northern pike were held in large tanks with flowing river water fed with external pumps at the experimental site until used. The more sensitive broad whitefish were contained in a pen located within the river itself until the airgun array was in position and then placed in a large tank on shore with flowing river water. Lake chub and YOY pike were held separately in smaller tanks of temperature controlled river water equipped with aeration and filtration in the DFO facility. All animals not used in experiments within 96 hours of capture were released. This study was approved by the Fisheries and Oceans Canada Animal Care Committee.

Measures of hearing

Hearing capabilities were measured using the auditory brainstem response (ABR), a noninvasive method of measuring the whole brain response to auditory stimuli (Corwin *et al.*, 1982; Kenyon *et al.*, 1998; Mann *et al.*, 2001; Scholik and Yan 2001, 2002; Smith *et al.*, 2004a, 2004b). ABR allows for a rapid assessment of hearing (15–20 minutes per fish) without training, and so it is possible to ascertain hearing loss very soon after exposure to sound.

Experiments were conducted with fish restrained in a mesh sling and suspended underwater in a large plastic cooler containing aerated river water (aeration was not used during testing). Lake chub and YOY northern pike were suspended so that the top of the head was approximately 9 cm below the surface of the water and 40 cm away from the underwater speaker (Aqua Synthesis). Because of their large size (350–670 mm), adult northern pike and broad whitefish were suspended so that they were at the same depth but 30 cm from the speaker. Water temperature ranged from 18 °C–20 °C.

A stainless steel recording electrode (Rochester Electro-Medical Inc., Tampa, FL) was inserted subdermally into the medial dorsal surface of the head over the brainstem while a similar reference electrode was placed into the dorsal midline surface of the fish near the anterior insertion of the dorsal fin. A ground electrode was placed in the water near the body of the fish. All exposed surfaces of the electrode tips that were not in direct contact with the fish were coated with enamel for insulation.

Sound stimuli were presented and ABR waveforms collected using a Tucker-Davis Technologies (TDT) physiology apparatus using SigGen and BioSig software (Tucker-Davis Technologies Inc., Gainesville, FL). Sounds were computer-

generated using TDT software and passed through a power amplifier (Hafler P1000) connected to the underwater speaker (Aqua Synthesis). Tone bursts were 50 ms in total duration and were gated with a Hanning window (similar to the conditions of past ABR studies in our laboratory; e.g., Mann *et al.*, 2001; Higgs *et al.*, 2001). Responses to each tone burst at each SPL were collected using the BioSig software package, with up to 1000 responses averaged for each stimulus frequency and level combination. In order to speed testing, if an evoked potential was obvious before 1000 averages was reached, the program was advanced to the next test condition. Sounds were presented 17.5 times per second. The SPLs of each presented frequency were confirmed using a calibrated underwater hydrophone (calibration sensitivity, -212 dB re 1 V/μPa; Reson TC 4013; 1 Hz–170 kHz response).

Sound intensity at each frequency was decreased in 6 dB steps until a stereotypical ABR was seen and then advanced to the next lower level (Fig. 1). Threshold was defined as the lowest level at which a response could be seen in the Fourier transform of the evoked potential that was 3 dB above background noise.

ABRs were determined for experimental, control, and baseline animals. Baseline animals were from the collected group but they were not placed in the experimental pen, and they thus served as controls for handling. The control group consisted of animals that were placed into the experimental pen and lowered to experimental depth and kept there for a period of time equivalent to the insonification period of experimental animals but without the sound exposure. Control and baseline results were similar, but baseline data are not reported here since they will be presented in a comparative study of hearing in a wide range of Mackenzie Delta species (Mann *et al.*, unpublished).

Once fish were tested with ABR they were deeply anesthetized with buffered MS-222 (an anesthetic for cold-blooded vertebrates) and then weighed and measured. The fish were then prepared for electron microscopic analysis to determine any effects on inner ear tissues (Popper *et al.*, unpublished).

Exposure paradigm

Broad whitefish and adult northern pike were placed into a 1 cubic meter holding pen made of 6.4 mm (1/4 inch) seine netting on a frame made of 12.8 mm (1/2 inch) metal rebar (top and bottom) and 12.8 mm (1/2 inch) lead-line (sides). The top of the pen had a hinged and latched wooden lid to allow access to the fish. The cage was set at the desired depth using floatation above and anchors below at the end of a fixed dock. Lake chub and YOY pike were exposed by placing them in a galvanized Gee minnow trap, the entrances to which were sealed with plastic netting. The experiments were conducted in 1.9 m of water with the pen and trap submerged so that they were centered about 1 m below the surface. Different fish species and life stages were exposed separately. For each test the fish were placed in the river and the airgun array was fired either five or 20 times.

Following sound exposure, the pen was lifted so that the top was just above the surface of the water. The fish were captured with a wetted net and placed into a holding tank with fresh river water. They were then taken by truck to the Fisheries and Oceans Canada lab (about a 90 second drive) where they were placed into aerated and filtered holding tanks of temperature controlled river water until they were used for ABR. With the exception of fish that were intentionally held for up to 24 hours post-exposure to look at recovery, all fish had ABRs measured within 1.5 hours of airgun exposure.

Seismic airgun array

The airgun array was a clustered array of eight equally spaced (70 cm between guns) SGI and SGII type sleeve guns with a total volume of 730 in.^{3} (12,000 cc) and with a total array dimension of 2.6 m in length and 1.22 m across. The volume of individual guns in the array ranged from 70 in.^{3} (1150 cc) to 150 in.^{3} (2460 cc). The airguns were deployed at 1.8 meters depth, were charged from a single air compressor and were fired with approximately 1900 psi (13.1 kPa) chamber pressure. The airgun array was fired manually and this led to small variations in the firing pressure of the airguns. The observed variability in the received level at the fish cages (see Table I) may be attributed to the manual firing of the airgun array.

The airgun array broadside was pointed toward the fish cage so that its maximum lateral pressure was radiated in the direction of the cage. The airgun array was positioned so that the cages were in the far field of the signal (i.e., where the pressure wave forms from the airgun added constructively). The position of the airgun array shifted slightly during 28 July, the first day of testing, due to high winds at the study site and had to be repositioned on 29 July, the second day of testing. The airgun array was 17 m from the fish cages on the first day of testing and 13 m on the second day of testing.

Airgun calibration

Both acoustic pressure and the acoustic particle velocity were measured directly adjacent to the fish cage during the exposure tests. Received levels inside the cage were not expected to be significantly different than those measured outside the cage. Measurements performed by JASCO Research Ltd. for a prior study, using similar mesh cages, indicated that airgun levels were not measurably different inside and outside of the cages (MacGillivray *et al.*, 2002).

The particle velocity was computed from differential measurements of the acoustic pressure using the pressure gradient method (Fahy, 1977). To do this, JASCO Research Ltd. designed and built an apparatus that consisted of a copper frame constructed so as to describe the three perpendicular Cartesian axes (x, y, and z). The distance from the origin point to each of the axis-ends was 50 cm. Four calibrated Reson TC4043 hydrophones (nominal sensitivity -201 dB re V/μPa) were mounted at the axis-ends while a single calibrated Reson TC4034 reference hydrophone (nominal sensitivity -218 dB re 1 V/μPa) was placed midway along the z axis of the pressure gradient sensor. A JASCO Research Ltd. UWINSTRU depth/attitude monitoring sensor mounted

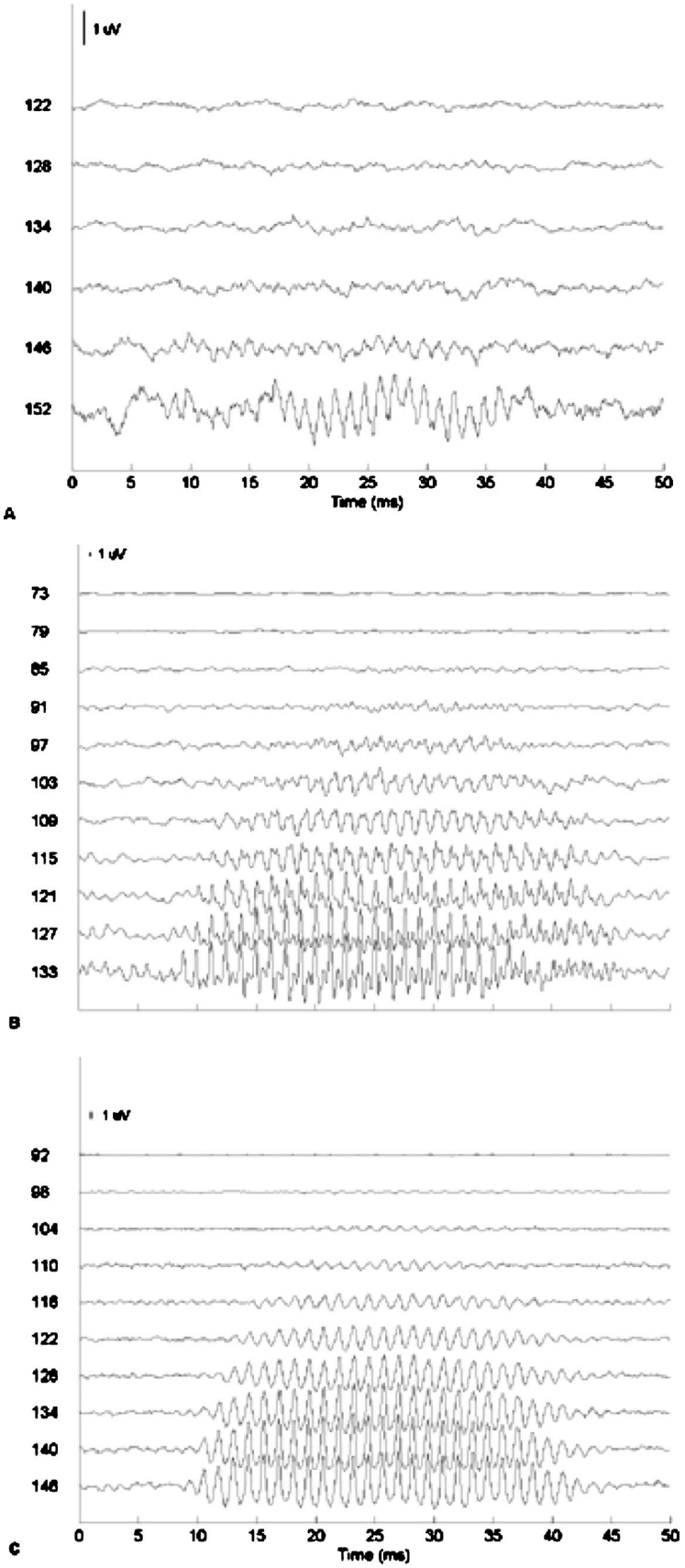


FIG. 1. Representative auditory brainstem response traces for a 400 Hz sound. The sound level (dB re 1 μPa) used for each trace is shown on the y-axis. The amplitude of the evoked potential is indicated by 1 μV bar shown at the top of each plot. Note that the scale is different for each plot. (A) Broad whitefish, (B) lake chub, and (C) adult northern pike.

along the x-axis of the frame to establish orientation of the apparatus for calculations of pressure and particle displacement.

The signals from the five hydrophones and the UWINSTRU were fed through custom shielded underwater cables to a laptop based acquisition system. The hydrophone signals were digitized using a Quatech DAQP-16 PCMCIA acquisition card using custom software and recorded to hard

TABLE I. Calibration data for each airgun shot and mean data for all shots. Tests 1–5 were on 7/28/2004. Tests 6–10 were on 7/29/2004. The various metrics used are discussed in the text.

Test	Species	Mean peak SPL (dB re 1 μPa)	Mean 90% RMS SPL (dB re 1 μPa)	Mean SEL (dB re 1 μPa$^2\cdot$s)	Mean peak velocity level (dB re 1 nm/s)	Mean 1 s RMS velocity level (dB re nm/s)	Mean peak intensity level (dB re 1 μPa)	Mean 1 s RMS intensity level (dB re 1 μPa)
1	Broad whitefish	209.5	199.4	180	139.6	112.6	202.6	176.5
2	Adult northern pike	207.3	197.7	178.3	139	111.7	202.0	175.2
3	Adult northern pike	207.5	198	178.3	139.4	112.1	202.1	175.5
4	Broad whitefish	209.9	199.8	179.7	140.7	113.3	203.4	176.8
5	Lake chub	205.2	195.1	175.9	136.7	109.7	200.1	173.1
6	Lake chub	206.8	197.9	177.9	140.2	112.1	202.4	175.6
7	Lake chub	205.7	196.2	176.7	136.9	110.1	199.5	173.7
8	Lake chub	207.1	197.4	177.5	138.5	111.2	201.0	174.5
9	YOY northern pike	207.5	197	177	139.5	110.4	201.8	173.6
10	YOY northern pike	206.2	195.3	175.9	136.7	108.8	199.1	171.6
Average		207.3	197.4	177.7	138.7	111.2	202.6	176.5
MIN		205.2	195.1	175.9	136.7	108.8	202.0	175.2
MAX		209.9	199.8	180	140.7	113.3	202.1	175.5

disk. The digital sampling rate for the acoustic signals was 20 kHz on each channel with 16-bit resolution (± 10 V maximum range). In addition, the signal from the single TC4034 reference hydrophone was amplified using an Ithaco 451M programmable gain amplifier. The TC4043 hydrophones have built-in preamplifiers and were not amplified prior to digitizing. The orientation and depth measured by the UWINSTRU were communicated via serial interface to the laptop and recorded into the logbook.

Acoustic metrics

For each exposure test, average received sound levels are reported using three standard metrics for periodic transient sources (as described in Richardson *et al.*, 1995): peak sound pressure level (Peak SPL), 90% RMS sound pressure level (90% RMS SPL), and sound exposure level (SEL) (see Table I). Sound pressure levels are reported in dB re 1 μPa and sound exposure levels in dB re μPa$\cdot$s^2. Acoustic pressure was measured in the frequency band 2 Hz–10 kHz.

The average acoustic particle velocity was measured for each exposure and is reported using two metrics, peak velocity level and 1 second RMS velocity level. Particle velocity levels are reported in dB re nm/s (the ANSI standard acoustic reference velocity). The finite baseline of the differential pressure measurement placed an upper limit on the maximum frequency at which particle velocity could accurately be measured. Thus, particle velocity was low-pass filtered at 1325 Hz, which corresponds to the 3 dB point of the estimated error of the velocity measurement. In addition, particle velocity was high-pass filtered at 150 Hz to reject low-frequency noise.

Acoustic intensity was computed from the product of the pressure and velocity traces. For each exposure, intensity is reported using two metrics, peak intensity level and 1 second RMS intensity level. Intensity level measurements are reported in dB re 0.676×10^{-18} W/m^2, the intensity of a 1 μPa plane wave.

Calibration of airgun array

A total of 10 noise exposure tests were performed on 28 and 29 July 2004 (Table I). Figure 2(A) shows a representative pressure waveform and its associated frequency spectrum as measured at the fish cages during the noise exposure tests. Figure 2(B) shows the velocity amplitude trace for the same shot, along with the frequency spectrum of the three axial velocity traces. Calibration data are summarized in Table I for each test. The calibration results for each test did not differ substantially from the average of all of the tests.

The ABR tank was calibrated by measuring the pressure gradient at each location. While experiments were conducted using a measure of acoustic pressure, particle motion calibrations were also calculated from the pressure gradients. Table

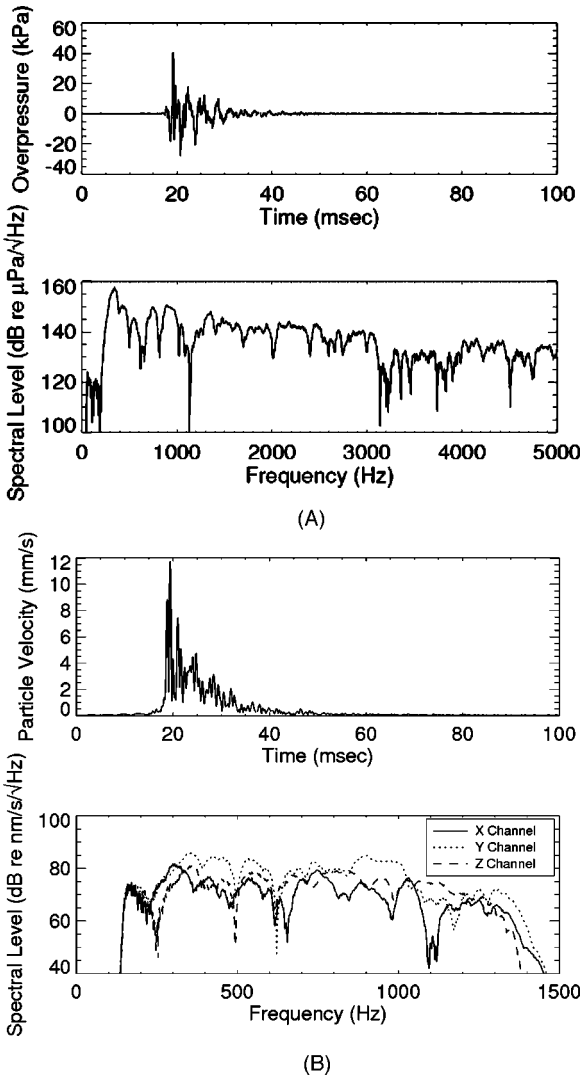


FIG. 2. (A) Acoustic pressure wave form and spectral levels for a single airgun array shot as measured during the noise exposure tests. (B) Particle velocity amplitude (top) and single-channel velocity spectral levels (bottom) as measured for a single airgun array shot during the noise exposure tests.

II shows the relationship between pressure and particle motion for each frequency at the head of the fish. There is a clear indication of a strong particle motion signal at each frequency tested. However, since it is not known whether the three species were detecting particle motion, pressure, or a combination of both signals, it is impossible to present hearing data in terms of which signals are most relevant to each species. At the same time, since there is a strong particle motion component at each frequency tested, any threshold shifts encountered most likely reflect a loss of detection capabilities of fishes to the signal(s) most relevant to their own hearing.

TABLE II. Calibration data for test tank showing particle velocity magnitudes corresponding to a 100 dB re 1 μPa sound pressure.

Frequency	Lake chub particle velocity (dB re 1 nm/s)	Pike/Whitefish (dB re 1 nm/s)
100	87	72
200	61	58
400	72	67
800	53	56
1600	24	12

Statistical analysis

The effect of exposure to seismic airgun noise and recovery from the exposure on auditory threshold levels were tested using separate ANOVAs with treatment (control or noise-exposed) and frequency as factors. Tukey's *post-hoc* test was used to make pairwise comparisons between treatments at specific frequencies when significant main effects were found with the overall ANOVA (Zar, 1998). SYSTAT (version 10) was used for all statistical tests.

RESULTS

We first provide general observations of hearing thresholds for each species examined (baseline data are presented in Mann *et al.*, unpublished) and then describe the effect of exposure to seismic airguns for each species separately. It should be noted that since we were collecting animals in the wild, we had little control over fish size. Analysis of size data relative to hearing thresholds, however, showed no difference in hearing sensitivity between fishes of the same species within the size ranges used (Mann *et al.*, unpublished). Consequently, data for all animals in each experimental and control group were pooled in presenting results. The only data not pooled were those for adult and YOY northern pike. While their hearing thresholds were similar between these two groups, the size differences in the specimens used were so great that it was decided to keep the groups separate in all data analyses.

Control hearing thresholds showed that lake chub has far more sensitive hearing and broader bandwidth of hearing than broad whitefish or northern pike (Figs. 3, 4, 6). This is as expected since the lake chub is a member of the superorder Otophysi, a group of hearing specialists that have a set of bones, the Weberian ossicles, which acoustically couple the swim bladder to the saccule of the inner ear. Hearing sensitivity is greater in the northern pike than the broad whitefish.

Our initial analysis of hearing sensitivity in the northern pike and broad whitefish showed that they could detect sounds up to 1600 Hz (the highest frequency used in this study). However, both species had very poor hearing at 800 and 1600 Hz. Since the speaker used in the ABR studies could not produce amplitudes that would be much higher than normal thresholds at 800 and 1600 Hz, it was not possible to examine hearing loss at these frequencies since we would have had to generate signals well above threshold to evaluate hearing loss (a higher hearing sensitivity). Thus, we chose to only test for threshold shift at 100, 200, and 400 Hz for these two species. Threshold shift was measured to 1600 Hz for the lake chub since its hearing bandwidth is greater and its normal thresholds are well below those of the other species.

General observations

While we did not do a standard necropsy or histopathology on test animals, general examination of the external

TABLE III. Hearing thresholds for broad whitefish controls and experimental animals exposed to five seismic shots.

Frequency	Mean	SD	SE	*N*	Frequency	Mean	SD	SE	*N*
	Control					Experimental: 5 shots, tested shortly after exposure			
100	114.8	13.99	5.29	7	100	115.7	5.02	2.24	5
200	112.8	9.47	3.58	7	200	108.5	5.02	2.24	5
400	113.1	18.31	6.92	7	400	109.7	11.54	5.16	5

anatomy post-exposure did not show any apparent effect of noise exposure as compared to controls. In addition, we did not note any bleeding or other overt effects on the eyes, gills, or internal organs in experimental or control groups when they were dissected in preparation for preservation of ear tissue for later analysis (Popper *et al.*, unpublished). The swim bladders were fully intact and inflated in all experimental and control specimens of all three species. Moreover, fish swam normally post-exposure and all fish that we maintained for use 24 hours post-exposure survived with no apparent adverse effects.

Broad whitefish

Fish ranged in size from 350 to 510 mm in standard length and 735 to 2810 grams in weight. Hearing was only measured immediately post-exposure due to difficulties in keeping these very sensitive animals alive in captivity. Hearing thresholds were obtained from five experimental fish and seven controls (Table III). Thresholds of whitefish exposed to five airgun shots were not significantly different from those of controls ($F=0.31$, $P=0.58$; Fig. 3), indicating that the airguns had no apparent effect on hearing in this species. There were no exposure effects on mortality.

Northern pike

We measured two groups of northern pike. One included adults from 360 to 670 mm in standard length and 430 to 2460 grams in weight. The second group included YOY fish that ranged from 70 to 110 mm in standard length and 1.7 to 8.8 grams in weight.

Adult pike exposed to five airgun shots exhibited mean thresholds that were higher than controls ($F=36.31$, $P<0.0001$), although this was significant only at 400 Hz ($P=0.0003$) because of low power ($N=4$ or 7; power $=80\%$, 66%, and 99% for 100, 200, and 400 Hz, respectively; Fig 4(A), Table IV). We plotted the threshold shift (exposed-control thresholds) to visualize trends in hearing loss across frequencies [Fig. 4(B)]. The greatest threshold shift was approximately 20 dB at 400 Hz. However, 18 hours after exposure to the airguns, the thresholds for northern pike were no longer significantly different from controls ($P>0.60$ for all frequencies), indicating complete recovery from hearing loss [Fig. 4(A)]. In contrast to the threshold shifts exhibited by adult pike, juvenile pike exhibited no hearing loss after being exposed to either five or 20 airgun shots ($P>0.10$, Fig. 5, Table V).

Lake chub

The response of lake chub (Table VI) was tested for both five and 20 airgun shots (Fig. 6) and for recovery from both signals (Fig. 7). Fish tested shortly after exposure to five airgun shots showed statistically significant threshold shifts at 200 ($P<0.0001$), 400 ($P=0.018$), and 1600 Hz ($P=0.001$). Lake chub that received 20 shots and then tested shortly after exposure showed hearing thresholds that were statistically different from controls ($P\leqslant 0.001$ for all frequencies). There was a mean difference in thresholds between animals exposed to five shots compared to those exposed to 20 shots ($F=57.08$, $P<0.0001$), but these differences were only significant at 400 Hz ($P<0.0001$) and 800 Hz [$P=0.027$; Fig. 6(A)]. The greatest threshold shifts

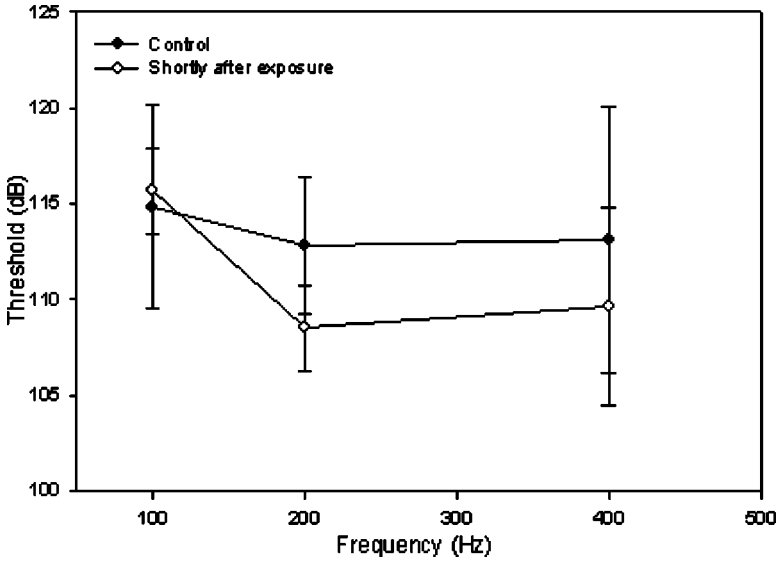


FIG. 3. Broad whitefish experimental vs. control data (mean ± SE) for fish exposed to five airgun shots.

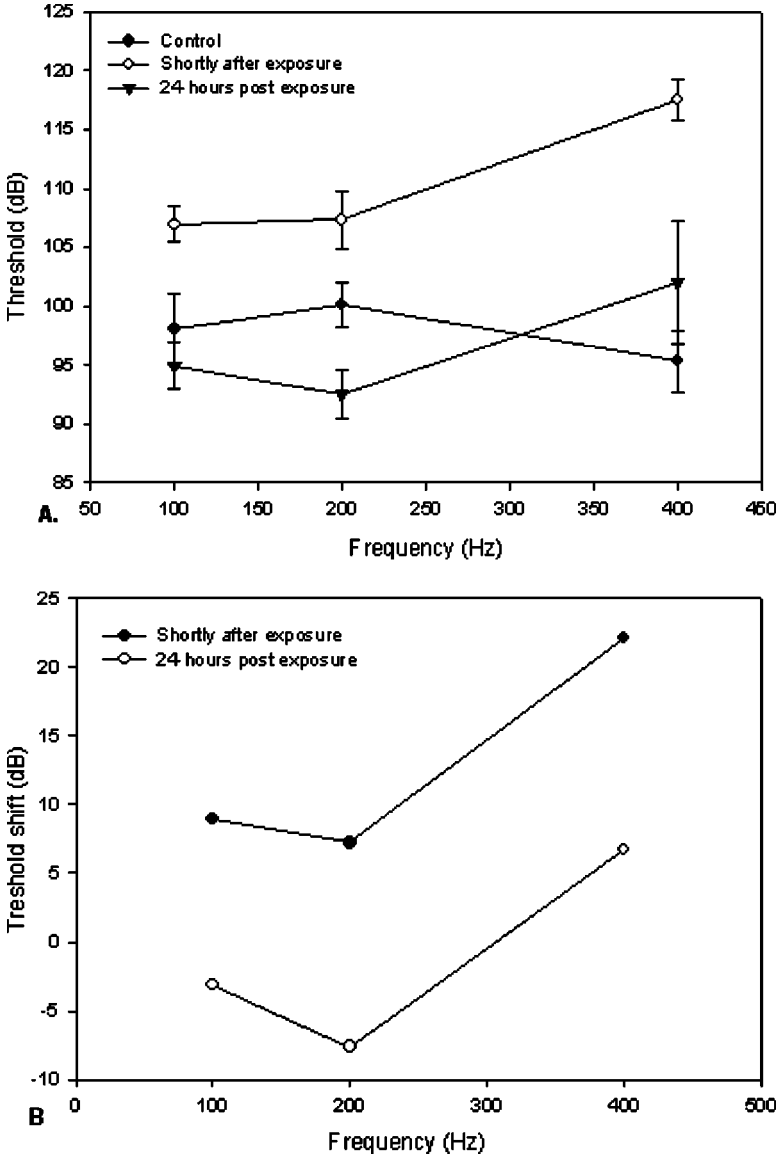


FIG. 4. (A) Thresholds for adult northern pike for controls and those exposed to five airgun shot immediately and 24-hours post-exposure. (B) Data for adult northern pike expressed as experimental threshold minus control data to show the effects of airgun exposure. Positive values indicate a hearing loss.

to five shots and 20 shots were approximately 25 dB at 200 Hz and 35 dB at 400 Hz, respectively [Fig. 6(B)]. Lake chubs tested 18 hours after exposure to five airgun shots had thresholds close to those of controls held for 18 hours (although this was not tested statistically because of low sample size) [Fig. 7(A)]. Chubs exposed to 20 airgun shots showed full recovery from hearing loss and had thresholds that were not significantly different from controls [$F = 0.18$, $P = 0.67$; Fig. 7(B)].

DISCUSSION

This study represents the first physiological evaluation of the effects of airgun use on the hearing sensitivity of fish. Earlier studies evaluating the effects of continuous noise and pure tones on hearing have shown temporary threshold shift (TTS) (e.g., Popper and Clarke, 1976; Scholik and Yan, 2001, 2002; Smith *et al.*, 2004a, 2004b). Such studies led to the concern that exposure to very intense sounds, such as

TABLE IV. Hearing thresholds for adult northern pike controls and experimental animals exposed to five seismic shots tested immediately post-exposure and another group tested 24 hours post-exposure.

Frequency	Mean	SD	SE	*N*	Frequency	Mean	SD	SE	*N*
Control					Experimental: Five shots, tested shortly post-exposure				
100	98.0	7.78	2.94	7	100	107.0	3.00	1.50	4
200	100.1	4.95	1.87	7	200	107.3	4.90	2.45	4
400	95.3	6.85	2.59	7	400	117.5	3.46	1.73	4
					Experimental: Five shots, tested 24 hours post-exposure				
					100	94.9	3.46	2.00	3
					200	92.5	3.46	2.00	3
					400	102.0	9.17	5.29	3

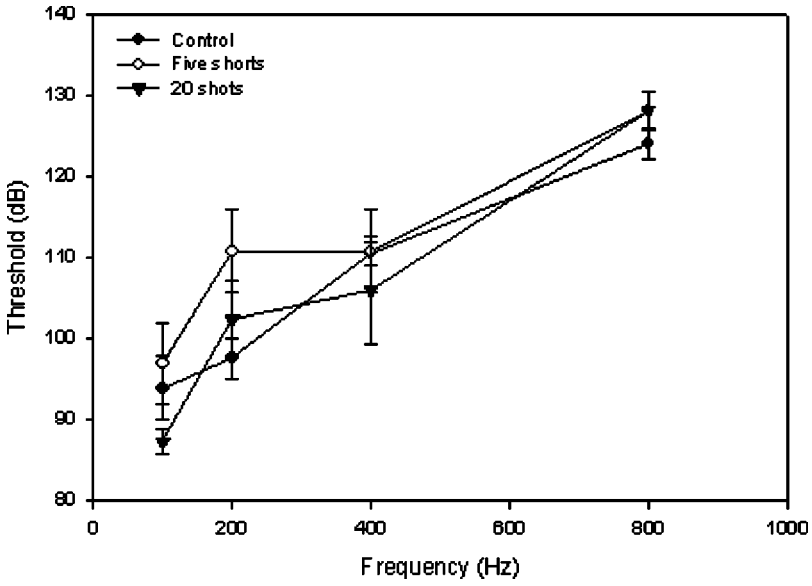


FIG. 5. Response of YOY northern pike to five and 20 airgun shots as compared to controls (mean ± SE). There is no statistical difference between exposed and control animals, indicating that airgun exposure did not affect hearing in these animals.

those from airguns, could impair fish hearing. At the same time, the sounds from an airgun are strikingly different from those used in earlier studies in several respects. Perhaps most importantly, unlike the sounds used in earlier studies, the sound from the airguns have an extremely rapid onset, contain energy over a broad frequency range, and have a much higher peak sound level (Fig. 2) than the noise or pure tones used in other studies. Thus, the sounds of airguns are closer to those of pile driving and explosions than sounds of ship noise or sonar.

The results of three fish species from the Mackenzie Delta demonstrate that there are substantial differences in the effects of airguns on the hearing thresholds of different species. Interestingly, the effects appear to have a correlation with hearing sensitivity of the fish. Thus, the broad whitefish, the species with poorest hearing sensitivity as measured in our apparatus, showed no apparent effects from the airgun exposure (five shots), while the lake chub, the species with most sensitive hearing, showed the most effect to both five and 20 airgun shots. The northern pike has hearing sensitivity that is between the two other species (albeit closer to that of the whitefish) and adult pike showed statistically significant hearing loss but less than that encountered with the chub. For reasons that we do not understand, however, YOY northern pike did not show any statistically significant hearing loss as a result of airgun exposure, even to 20 shots, although this may be related to issues associated with development of the auditory system (e.g., Kenyon *et al.*, 1998; Wysocki and Ladich, 2001).

The sound spectrum of the airgun array shots (Fig. 2) in the riverine exposure showed most energy above 300 Hz. While the spectrum of airguns can be expected to contain a large low frequency component (e.g., McCauley *et al.*, 2003), there is considerable loss of low frequency energy in shallow waters (e.g., Rogers and Cox, 1986), thereby resulting in a difference in the spectral components of airguns events in waters of different depths. Significantly, the spectrum of the airguns used in this study generally matched the threshold shifts observed in adult northern pike and in lake chub. Both of these species showed the greatest threshold shifts at 400 Hz, with less hearing loss at 100 Hz and 200 Hz. However, the lake chub did show large threshold shifts at 200 Hz, even though there was relatively little energy at 200 Hz in the shot.

There are suggestions in the literature that the effects of high intensity sound on the hearing abilities of fish are related to the level of the stimulus sound above the threshold of the fish (Hastings *et al.*, 1996; Smith *et al.*, 2004a, 2004b). It has been hypothesized that noise-induced threshold shifts in fish are linearly related to the sound pressure

TABLE V. Hearing thresholds for YOY northern pike controls and experimental animals exposed to five seismic shots and 20 seismic shots.

Frequency	Mean	SD	SE	*N*	Frequency	Mean	SD	SE	*N*
YOY pike: controls					YOY pike: five shot tested shortly post-exposure				
100	93.9	7.75	3.87	4	100	96.9	11.22	5.02	5
200	97.6	4.90	2.45	4	200	110.8	11.54	5.16	5
400	110.5	3.00	1.50	4	400	110.8	11.54	5.16	5
800	124.0	3.46	1.73	4	800	128.2	5.02	2.24	5
					YOY pike: 20 shots, tested shortly after exposure				
					100	87.3	3.29	1.47	5
					200	102.4	10.73	4.80	5
					400	106.0	14.70	6.57	5
					800	128.2	5.02	2.24	5

TABLE VI. Hearing threshold for lake chub controls tested immediately and controls tested after 18 hours and for experimental animals exposed to five and 20 shots tested immediately and 18 hours post-exposure.

Frequency	Mean	SD	SE	*N*	Frequency	Mean	SD	SE	*N*
Control					Experimental: 5 shots, test shortly post-exposure				
100	80.4	5.20	2.33	5	100	94.1	5.02	2.24	5
200	73	3.82	1.71	5	200	97.7	7.82	3.50	5
400	70.9	10.29	4.60	5	400	87.2	7.82	3.50	5
800	70.4	6.68	2.99	5	800	90.9	7.82	3.50	5
1600	89.7	7.63	3.41	5	1600	101.1	4.24	1.90	5
Control: fish kept 18–24 h post-exposure					Experimental: 5 shots, tested 18 h post-exposure				
100	87.6	7.77	3.17	6	100	89.9	4.24	3.00	2
200	79.2	10.50	4.29	6	200	87.5	4.24	3.00	2
400	77.3	4.33	1.77	6	400	83.0	4.24	3.00	2
800	82.5	6.93	2.83	6	800	83.7	0.00	0.00	2
1600	92.6	6.99	2.85	6	1600	89.1	8.49	6.00	2
					Experimental: 20 shots, tested shortly post-exposure				
					100	100.4	3.00	1.50	4
					200	110.0	3.00	1.50	4
					400	111.5	5.74	2.87	4
					800	104.7	3.46	1.73	4
					1600	108.6	3.00	1.50	4
					Experimental: 20 shots, tested 18 h post-exposure				
					100	93.5	7.63	2.89	7
					200	77.0	7.63	2.89	7
					400	77.7	7.52	2.84	7
					800	80.7	11.86	4.48	7
					1600	94.8	8.38	3.17	7

difference (SPD) between the sound pressure of the noise and the baseline hearing threshold of the fish (called the linear threshold shift, or LINTS hypothesis; Smith *et al.*, 2004b), as has been found in birds and mammals. Since the baseline thresholds of fish vary with frequency, this difference is calculated separately for each frequency tested. The linear TTS relationships of Smith *et al.* (2004b) were found after exposing fish to continuous sounds. In order to examine if this LINTS relationship is valid for more impulsive, short duration sounds such as our seismic airgun stimulus, we plotted our TTS data against SPD. The sound level used to calculate SPD was the measured 90% RMS SPL (dB re 1 μPa) of the airgun shots (see Table I). Regression analysis was used to examine the relationship between SPD and TTS.

Lake chub exposed to five and 20 airgun shots both exhibited a significant linear relationship between SPD above baseline hearing thresholds and TTS ($P<0.0001$; Fig. 8). Lake chub that were exposed to 20 airgun shots had TTSs that were greater than those exposed to only five airgun shots (see Results), and their LINTS relationship had a slightly greater slope as shown by a significant SPD and shot treatment interaction ($F=3.53$, $P=0.016$; Fig. 8).

The LINTS relationship is more evident for lake chub, which are hearing specialists with lower baseline hearing thresholds, than for pike and whitefish, which are hearing generalists with higher baseline hearing thresholds. As a result of the differences in hearing capabilities between these two groups, the SPD of our airgun source above baseline hearing thresholds is generally greater for chubs and minimal in pike and whitefish. Despite this fact and the low number of frequencies tested, adult pike exhibited a significant LINTS relationship (TTS$=0.53$, SPD$=45.50$; $R^2=0.21$, $P=0.040$), although whitefish did not ($P=0.067$). When all three species are plotted simultaneously, a significant LINTS relationship exists (Fig. 9).

This finding supports the LINTS hypothesis and suggests that such a relationship is valid for TTS induced by both continuous and impulsive sound sources. The result is a predictable linear relationship between SPD above baseline thresholds and TTS for these three species with varying hearing capabilities (Fig. 9). These findings support the idea that the LINTS relationship may ultimately be usable by fisheries managers attempting to mitigate the effects of intense anthropogenic sounds on fishes. In general, the only information that fisheries managers would need to predict such relationships is the audiogram of the species of interest and the sound spectrum, level, and duration of the sound stimulus in question.

While we did not attempt to use different levels of airgun sounds, we were able to use different total energy exposure in both the lake chub and northern pike, and this is equivalent to the effects of higher sound levels for noise studies if one assumes that hearing loss is a response to total energy impinging upon the animal, as predicted by the LINTS hypothesis. Our findings show that as the total energy of exposure increases (20 versus five airgun shots), there is a substantial increase in TTS in lake chub but not in northern pike YOY (which showed no TTS to any sounds).

Clearly, it would have been useful to have tested the LINTS idea with broad whitefish, as the LINTS hypothesis and the results with lake chub and northern pike suggest that

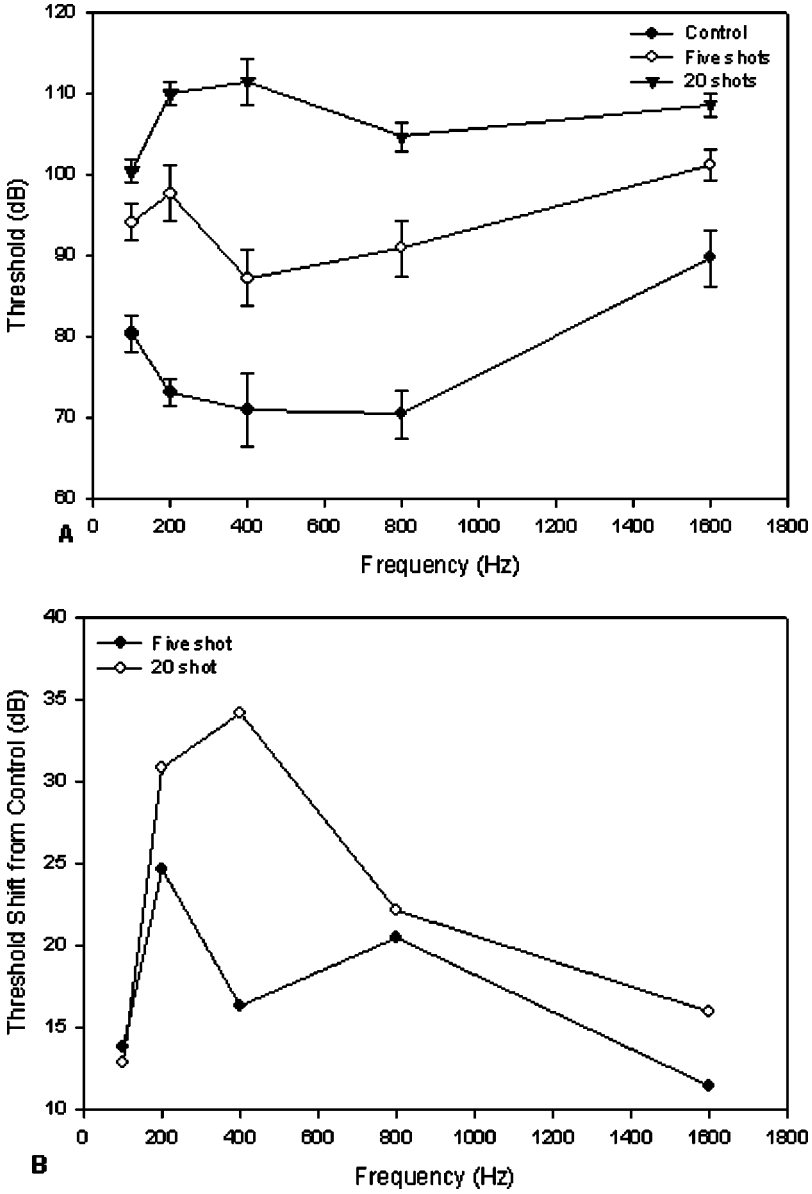


FIG. 6. Noise exposure data for lake chub exposed to five and 20 airgun shots. (A) Threshold data for controls and animals tested immediately after exposure. (B) Data expressed in terms of difference between exposure and threshold hearing levels.

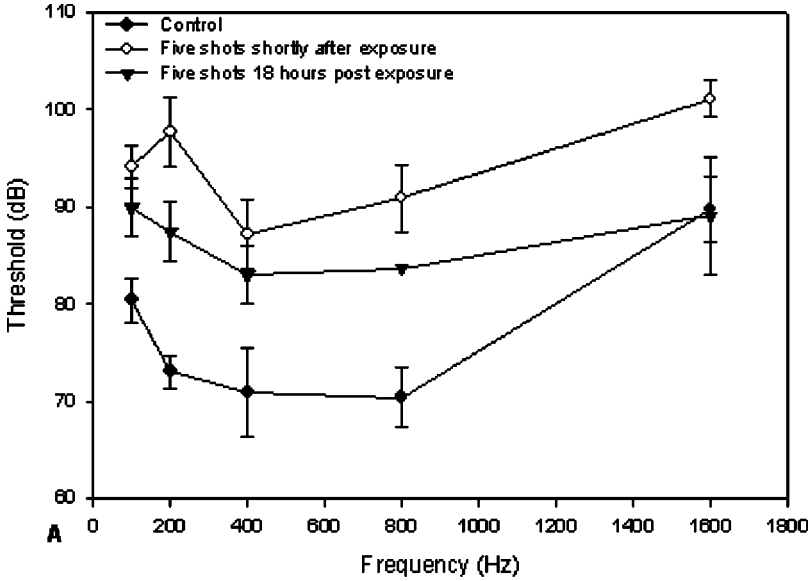
had whitefish been exposed to 20 airgun shots they might have demonstrated TTS in that species as well. However, this could not be done since insufficient specimens were available.

Despite the presence of a hearing loss immediately post-exposure, tests on lake chub and northern pike, 18 to 24 hours post-exposure, respectively, to five airgun shots shows a return to about normal thresholds. Thus, the hearing loss encountered after exposure appears to be temporary threshold shift, although any final assessment of whether there is longer term damage that shows up later post-exposure will have to await microscopic analysis of the ear tissue (Popper *et al.*, unpublished).

There were no obvious trends in the hearing measurements made from a group of exposed animals over the 1.5 hours it took to test them. It would be of considerable interest for future studies to examine the time course of recovery, but this would require the experimenters to be able to expose and test fish on a very tight time schedule—something not possible during the course of these experiments.

One concern is that our results show that exposure to 20 airgun shots is very likely to cause a TTS in every species exposed, except juvenile northern pike. It is unlikely that fish would encounter 20 shots in a river seismic survey unless they were ''herded'' with the survey vessel. Only strong swimming fish such as adult broad whitefish would be capable of being ''herded'' by the sound source and potentially stay in the vicinity of the airguns over multiple emissions. Thus the test with 20 shots over 15 minutes at 210 dB re 1 μPa probably represents an extreme exposure for fishes in the Mackenzie River Delta. The actual exposure of fish to a seismic survey depends on the speed of the survey vessel and the movements of the fish (Jorgensen *et al.*, unpublished). To understand the actual exposure will require behavioral studies on the movements of fishes in response to airgun surveys.

The results of this study have implications for airgun surveys, particularly in riverine situations. Our experimental paradigm was designed around a 2D survey that might be conducted in a river where the airgun vessel steadily moves in one direction, as opposed to off-shore 3D surveys where



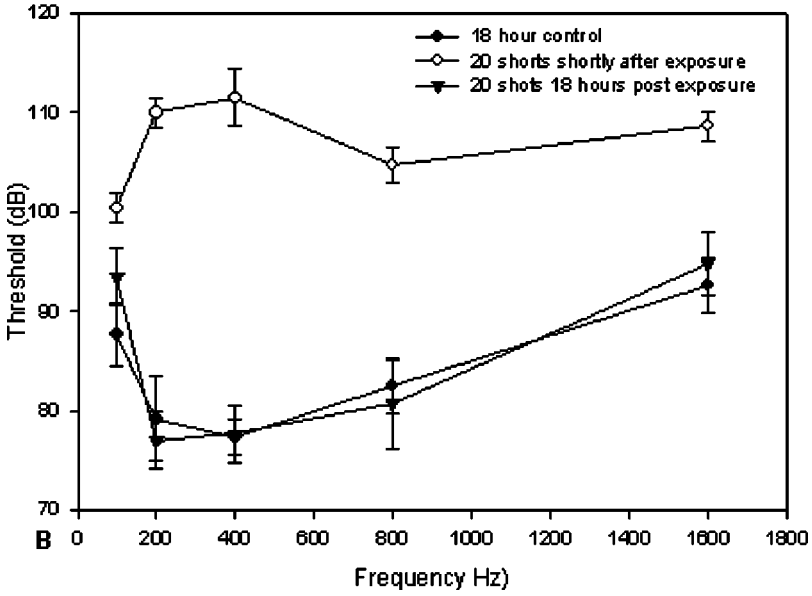


FIG. 7. Recovery of lake chub from exposure to five (A) and 20 (B) airgun shots.

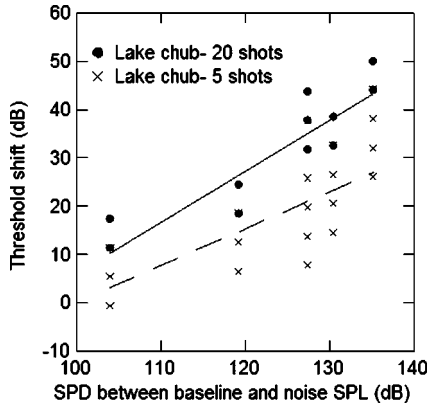


FIG. 8. Relationship between sound pressure difference (SPD) between the noise level and baseline hearing thresholds and temporary threshold shifts (TTS) for lake chub exposed to 5 or 20 shots of the airgun. The dashed and solid lines represent the linear regression relationships for chub exposed to 5 (TTS=0.76 SPD−75.84, $R^2=0.57$, $P<0.001$) and 20 airgun shots (TTS =1.06 SPD−99.44, $R^2=0.83$, $P<0.001$), respectively. Each data point represents the TTS ($N=4-5$) at each of the five frequencies tested.

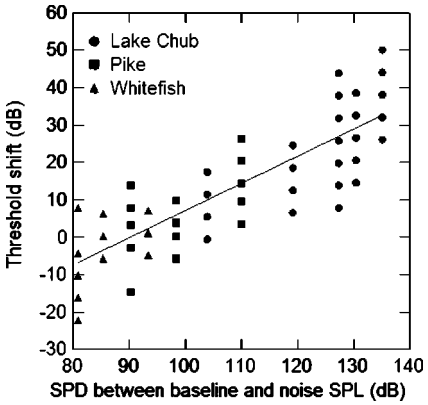


FIG. 9. Relationship between sound pressure difference (SPD) between the noise level and baseline hearing thresholds and temporary threshold shifts (TTS) of adult northern pike, lake chub, and broad whitefish exposed to the airgun. The solid line represents the linear regression relationship for all three species (TTS=0.53 SPD−48.40, $R^2=0.40$, $P<0.0001$). Each data point represents the mean TTS ($N=4-5$) at each of the frequencies tested.

the airguns are often towed back and fourth over parallel tracks (Bott, 1999). Fish in a river are exposed to airguns as the tow boat approaches and then passes by. In contrast, in a 3D off-shore seismic exploration program, a resident fish may be exposed to the airgun over and over again as the seismic array is moved across parallel paths.

In design of our study, we presented fish with five exposures to the airgun at 40 second intervals so that the fish were exposed to a steady sound level (Table I). In contrast, a normal survey might present signals as often as every 10 seconds (this could not be done in this study due to limitations of the compressor used to charge the guns). Since a seismic survey vessel is moving, a stationary fish subject is exposed to the maximum level once in a sequence of exposures. Moreover, the majority of exposed fishes during a seismic survey are likely to be at greater distances from the source than those in this study and thus receive a somewhat lower sound level. Though these factors do not compensate for the more frequent exposure in an actual survey, without use of a seismic vessel for experiments of this sort it is likely that our experiments presented fish with an approximate ''worst case'' with regard to seismic stimulation.

Based upon this being a ''worst case'' and the differences in effects on different species, it may be possible to suggest some general concepts with regard to seismic stimulation. First, it appears that it may be possible to predict whether a fish will show hearing loss as a consequence of exposure to airgun noise based upon baseline hearing thresholds of a particular species and using the LINTS determinations. Fish with poorer hearing in this study, such as pike, showed little hearing loss, while the fish with the best hearing, the lake chub, had the most loss.

Second, our data indicate recovery of hearing loss within 18 and 24 hours in the lake chub and northern pike, respectively, even after exposure to 20 airgun shots. It is unclear why complete recovery did not occur in the two chubs exposed to five airgun shots, and more data are needed to draw conclusions about recovery. However, despite recovery it is important to note that during the period of TTS fish may still be impaired in their ability to survive since they would have some loss of their ability to hear biologically relevant sounds.

Comparison with other airgun studies

The only other experimental airgun study that looked directly at the effects on fish physiology was done by McCauley *et al.* (2003) using a different paradigm and different species than used here. Thus, a direct comparison between the two studies is tenuous at best. The McCauley *et al.* study only looked at effects on ear tissue and did not examine the hearing ability of the experimental species, the pink snapper, and there are no data on hearing capabilities of that species. At the same time, McCauley *et al.* reported a profound and long-lasting effect on the sensory cells in the saccule of the ear, and it remains to be seen whether the same effect is found in this study (Popper *et al.*, unpublished).

There are several other differences in the two studies that make it hard to compare results. First, McCauley *et al.* (2003) used a 330 cc (20 $in.^3$) single airgun as compared to the 12 000 cc (720 $in.^3$) array used in this study. McCauley *et al.* also exposed fish with frequent emissions in two periods of almost 1 hour each to partially mimic the kinds of exposure fish might get were they stationary and exposed to an airgun that moves back and forth, as happens in 3D marine surveys. Thus, the cumulative exposure to the airgun in the McCauley *et al.* study may have been greater than in this study. Second, the McCauley *et al.* study was in deeper water and so the spectrum of the sounds to which the fish were exposed in the two studies differed somewhat, with there being more energy below 300 Hz in McCauley *et al.* than in the current investigation.

It clearly would be important to replicate the McCauley *et al.* (2003) study to ascertain effects on hearing, as well as on different fish species. Similarly, it would be of value to replicate our study using even greater stimulus levels or higher exposure rates in order to test whether the fish exposed to such sounds would ultimately show permanent hearing loss that would be associated with the loss of sensory cells found by McCauley *et al.* (2003).

Caveats and future studies

The results from this study provide a qualitative model to predict when exposure to airguns may have an effect on hearing capabilities of fishes. However, there are several caveats that must accompany our results before they are broadly applicable to other studies. We also suggest that additional studies need to be conducted to help resolve the remaining issues.

First, while we studied fishes with different hearing capabilities, since there is substantial diversity in the structure of the auditory systems of different species (e.g., Popper and Carlson, 1998; Popper *et al.*, 2003), it would be of use to look at the effects of airguns on species with other hearing specializations.

Second, while our results support the LINTS hypothesis in suggesting that increased total noise exposure will result in increased hearing loss, this idea needs to be tested more directly with airgun exposures of different total energy levels, including different numbers of airgun shots.

Third, it is not clear whether an increase in total noise exposure will result in permanent hearing loss (permanent threshold shift, PTS), and how the results reported here fit with the recent work of McCauley *et al.* (2003) that showed significant inner ear damage as a result of noise exposure that probably exceeded any used here. Future studies from our laboratory will evaluate ear structure in the Mackenzie River animals, but due to the nature of the species selected and our holding facilities at Inuvik, it was not possible to hold fish for the extended periods of time used by McCauley *et al.* It would be of considerable value to replicate our studies but then hold fish for many days and weeks post airgun exposure to determine if there is damage to the auditory system and if this is manifest in late onset hearing loss.

Fourth, our work was done in a river using an exposure paradigm designed to mimic a single pass exposure to a seismic source. In order to compare to the impacts of a 3D seismic survey, it will be important to replicate our type of

behavioral study using a sound exposure paradigm more like that used by McCauley *et al.* (2003) to mimic repeated exposure to sound.

Fifth, it must be kept in mind that this study was done in a river, with the fish in 1.9 m of water. Sound propagates differently in shallow water than it does in deep water, with much less propagation of low frequency energy in shallow water (Rogers and Cox, 1986). It is therefore important to determine if the effects of airgun sound on fish are the same in deep water as they are in shallow water.

Finally, we caution that the results reported here, while highly informative, are not the final word on the effects of airguns on fishes. And they are clearly *not* applicable to the potential effects from other anthropogenic sources such as shipping, sonar, or other more or less ''continuous'' signals or signals that do not have rapid onsets. Moreover, extrapolation to other species must still be done with considerable caution.

ACKNOWLEDGMENTS

This study was supported by Fisheries and Oceans Canada and the Department of Indian Affairs and Northern Development (DIAND) of the Canadian Government, the Program of Energy Research and Development (PERD), the Inuvialuit Fisheries Joint Management Committee (FJMC), and WesternGeco. The work would not have been possible without the cooperation and support of numerous people: Dave Tyson, Eric Gyselman, John Jorgensen, Marty Bergman, Ron Allen, and the DFO Inuvik office staff; Terry Jackson and his crew from Conquest Seismic Services Ltd.; Steve Whidden from WesternGeco; Les Harris from the Gwich'in Renewable Resource Board, Kevin Bill, Andrea Hoyt and the FJMC mentoring program students—Gerald Kisoun, Noel Cockney, Candice Cockney; Angus Alunik, and Keith Rosindell. The authors are grateful to Moe Grant for permitting us to use her property for this study and to Edward Dillon and Merik Allen for their valuable local knowledge and assistance in fish capture.

Amoser, S., and Ladich, F. (**2003**). ''Diversity in noise-induced temporary hearing loss in otophysine fishes,'' J. Acoust. Soc. Am. **113**, 2170–2179.

Amoser, S., Wyoscki, L. E., and Ladich, F. (**2004**). ''Noise emission during the first powerboat race in an Alpine lake and potential impact on fish communities,'' J. Acoust. Soc. Am. **116**, 3789–3797.

Bott, R. (**1999**). *Our Petroleum Challenge: Exploring Canada's Oil and Gas Industry*, 6th ed. (Petroleum Communication Foundation, Calgary, Canada).

Corwin, J. T., Bullock, T. H., and Schweitzer, J. (**1982**). ''The auditory brainstem response in five vertebrate classes,'' Electroencephalogr. Clin. Neurophysiol. **54**, 629–641.

Cott, P. A., Hanna, B. W., and Dahl, J. A. (**2003**). ''Discussion on seismic exploration in the Northwest Territories 2000–2003,'' Can. Spec. Publ. Fish. Aquat. Sci. **2648**, vi +36 p.

Enger, P. S. (**1981**). ''Frequency discrimination in teleosts—central or peripheral?'' in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 243–255.

Fahy, F. J. (**1977**). ''Measurement of acoustic intensity using the cross-spectral density of two microphone signals,'' J. Acoust. Soc. Am. **62**, 1057–1059.

Hastings, M. C., Popper, A. N., Finneran, J. J., and Lanford, P. J. (**1996**). ''Effect of low frequency underwater sound on hair cells of the inner ear and lateral line of the teleost fish *Astronotus ocellatus*,'' J. Acoust. Soc. Am. **99**, 1759–1766.

Higgs, D. M., Souza, M. J., Wilkins, H. R., Presson, J. C., and Popper, A. N. (**2001**). ''Age- and size-related changes in the inner ear and hearing ability of the adult zebrafish (*Danio rerio*),'' J. Otolaryngol. **3**, 174–184.

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.

MacGillivray, A., Austin, M., and Hannay, D. (**2002**). ''Acoustic level measurements of airgun arrays from WesternGeco's 2002 Mackenzie River seismic project,'' JASCO Research Ltd., Victoria, B. C., for WesternGeco Ltd., Calgary, AB. 26 pp.

Mann, D. A., Higgs, D. M., Tavolga, W. N., Souza, M. J., and Popper, A. N. (**2001**). ''Ultrasound detection by clupeiform fishes,'' J. Acoust. Soc. Am. **109**, 3048–3054.

McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). ''High intensity anthropogenic sound damages fish ears,'' J. Acoust. Soc. Am. **113**, 638–642.

Myrberg, A. A., Jr. (**1980**). ''Ocean noise and the behavior of marine animals,'' in *Advanced Concepts in Ocean Measurements for Marine Biology*, edited by F. P. Diemer, F. J. Vernberg, and D. V. Mirkes (University of South Carolina Press, Columbia), pp. 461–491.

NRC (National Research Council) (**2003**). *Ocean Noise and Marine Mammals* (National Academy Press, Washington, DC).

NRC (National Research Council) (**2005**). *Marine Mammal Populations and Ocean Noise: Determining when Noise Causes Biologically Significant Effeccts* (National Academy Press, Washington, DC).

Popper, A. N. (**2003**). ''Effects of anthropogenic sound on fishes,'' Fisheries **28**, 24–31.

Popper, A. N., and Carlson, T. J. (**1998**). ''Application of the use of sound to control fish behavior,'' Trans. Am. Fish. Soc. **127**, 673–707.

Popper, A. N., and Clarke, N. L. (**1976**). ''The auditory system of the goldfish (*Carassius auratus*): Effects of intense acoustic stimulation,'' Comp. Biochem. Physiol. A **53A**, 11–18.

Popper, A. N., Fay, R. R., Platt, C., and Sand, O. (**2003**). ''Sound detection mechanisms and capabilities of teleost fishes,'' in *Sensory Processing in Aquatic Environments*, edited by S. P. Collin and N. J. Marshall (Springer-Verlag, New York), pp. 3–38.

Popper, A. N., Fewtrell, J., Smith, M. E., and McCauley, R. D. (**2004**). ''Anthropogenic sound: Effects on the behavior and physiology of fishes,'' Mar. Technol. Soc. J. **37**, 35–40.

Richardson, W. J., Greene, C. R., Jr., Malme, C. L., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, New York).

Rogers, P. H., and Cox, M. (**1988**). ''Underwater sound as a biological stimulus,'' in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga (Springer-Verlag, New York), pp. 131–149.

Scholik, A. R., and Yan, H. Y. (**2001**). ''Effects of underwater noise on auditory sensitivity of a cyprinid fish,'' Hear. Res. **152**, 17–24.

Scholik, A. R., and Yan, H. Y. (**2002**). ''The effects of noise on the auditory sensitivity of the bluegill sunfish, *Lepomis macrochirus*,'' Comp. Biochem. Physiol. A **133A**, 43–52.

Smith, M. E., Kane, A. S., and Popper, A. N. (**2004a**). ''Noise-induced stress response and hearing loss in goldfish (*Carassius auratus*),'' J. Exp. Biol. **207**, 427–435.

Smith, M. E., Kane, A. S., and Popper, A. N. (**2004b**). ''Acoustical stress and hearing sensitivity in fishes: Does the linear threshold shift hypothesis hold water?'' J. Exp. Biol. **207**, 3591–3602.

Tallman, R. F., and Reist, J. D. (editors) (**1997**). ''The proceedings of the broad whitefish workshop: the biology, traditional knowledge and scientific management of broad whitefish (*Coregonus nasus* (Pallas)) in the lower Mackenzie River,'' Can. Spec. Publ. Fish. Aquat. Sci. **2193**, xi+219 p.

Wartzog, D., Popper, A. N., Gordon, J., and Merrill, J. (**2004**). ''Factors affecting the responses of marine mammals to acoustic disturbance,'' Mar. Technol. Soc. J. **37**, 6–15.

Wysocki, L. E., and Ladich, F. (**2001**). ''The ontogenetic development of auditory sensitivity, vocalization and acoustic communication in the labyrinth fish *Trichopsis vittata*,'' J. Comp. Physiol. A **187**, 177–187.

Zar, J. H. (**1998**). *Biostatistical Analysis*, 4th ed. (Prentice-Hall, Upper Saddle River, NJ).